#21 5/20/93

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	Group Art Unit: 1804
COMAI, ET AL.	Examiner: P. MOODY
Serial No.: 07/985,742	
Filed: December 4, 1992	DECLARATION OF MARGARET P. SANGER
For: FIGWORT PLANT PROMOTER AND USES	

Honorable Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

1. I have my Bachelors' (1976) and Master of Science (1979) degrees in biology from the California State University at Northridge. Before receiving a doctoral degree, I was involved in research biochemistry at Michigan State University. In 1987, I received a PhD from UC Davis in Plant Microbe Interactions. After receiving my doctoral degree, I performed research in molecular virology at Calgene and in other positions. At the present time, I am a Visiting Scholar in the Department of Plant Pathology and in the Department of Viticulture and Enology at UC Davis. I am one of the inventors named on the subject patent application.

I. INTRODUCTION

2. I have been asked to provide a declaration on the issue of whether or not the invention claimed in this application, namely the FMV 34S promoter, would have been obvious at the time that the invention was made. I submitted a previous declaration in connection with parent Application

Serial No. 07/431,429. The substance of the earlier declaration was that the invention claimed in this application was made at least prior to November 13, 1988. Therefore, in commenting on the question of whether the FMV 34S would have been obvious, I am speaking from the perspective of the state of the art in 1988 and prior to 1988.

3. In 1988, there was only one plant viral promoter system understood: CaMV 35S. Since this single system represented the state of the art, it would not have been possible to make any generalizations or predictions about other viral promoter systems. A single data point (CaMV 35S) could not reasonably provide a scientific basis for making sound predictions of likely promoter activity and strength. There was sufficient information to expect that there was a promoter site to examine. However, there was no information as to the strength and utility of said promoter. I found the promoter sequence (CAT and TATA without the upstream region) that was visible to be of almost zero activity and comparable to other examples of low activity sequences. The useful activity of this promoter derives from sequences distant from the simple promoter motif. There was no sequence identity between this promoter and the CaMV 35S promoter, so promoter strenth was not obvious. At that time, there was no general knowledge for caulimoviruses that sequences distant from this promoter would enhance its utility. We did not know what to expect.

4. Even now, about five years later, after the identification and characterization of many plant viral promoter systems, a review of promoter structure provides little or no guidance which would enable one to predict the function of a specific

structure. Ultimately, one has to isolate, implement and characterize a promoter to see if the promoter has the desired activity. Below, I describe in detail how function does not necessarily follow structure in comparing CaMV 35S to FMV 34S.

II. RELATION OF DNA SEQUENCE TO PROMOTER ACTIVITY

5. The notion that promoter quality can be predicted from sequence gazing is untenable. The sequences of the CaMV 35S and FMV 34S promoters are quite different, except for the basic CAAT and TATA motifs, which are common to most eucaryotic promoters. These basic motifs appear in the 19S promoter of CaMV, yet this promoter is much weaker than the 35S promoter (Guilley et al, 1982). The promoter in the region of SoyCMV gene III is similar to the 19S promoter from CaMV, yet it is comparable in strength in vitro to that of the 35S (Hasagawa et al, 1989). In fact, these motifs are well away from the interesting regions of the promoter which contribute to promoter activity. However, despite the great dissimilarities, the strengths of these promoters are comparable (Sanger et al, 1990).

6. Moreover, small, superficially insignificant changes in the overall context of the CaMV 35S promoter can have big effects on promoter strength (Ow et al, 1987, Odell et al, 1985).

7. Therefore, even though sequence analysis can reveal the extent of homology in related domains (e.g., the CAAT and TATA motifs), an investigation involving the isolation, purification, and engineering into a quantitative expression system is essential for any promoter before a meaningful conclusion about its relative strength can be made. We can now say that the 35S promoter from CaMV is comparable to the 34S promoter from FMV. Having said

that, we can infer that the parallel contexts of these promoter domains must therefore call for a strong promoter at that general position for any caulimovirus. But, before the characterization of the 34S FMV promoter was undertaken, such a claim would have been completely without substance, based on extrapolation from one single example.

III. RELATIONSHIP BETWEEN PROMOTER STRENGTH AND VIRAL TITER

8. There is no proven relationship between the strength of a given viral promoter and the aggressiveness of the parent virus in planta. Furthermore, the severity, viral titer, aggressiveness of CaMV is much different from that of FMV, so that if, e.g. viral titer were to be an indicator to promoter strength, we would expect the FMV 34S promoter to be much weaker than the CaMV 35S.

9. We have had great difficulty in the purification of virus from the strain of FMV from which the 34S promoter was derived. However, using the same procedure, we can readily purify mg amounts of CaMV, of the strain from which the 35S promoter was derived. This difference reflects the difference in their titers. The titer of CaMV, of the type from which the 35S promoter was derived, is approx. 6 to 10 mg virus recovered per 1 kg infected tissue (Hull et al, 1976). The titer of wild-type FMV, the type from which the 34S promoter was derived, is approx. 330 ug virus per kg tissue (Shepherd et al, 1987). The adaptation of FMV during the course of routine laboratory maintenance of viral stocks passaged repeatedly through a greenhouse host will produce a raised-titer strain. This adaptation is thought to be due to changes within viral gene VI (Shepherd et

al, 1987; Sanger et al, 1991); it has not been suggested that the change derives from the 34S promoter. The 34S promoter characterized by Sanger et al, (1990) is from a low titer strain of FMV, yet it is comparable in strength to the 35S promoter. We would be very much surprised if the strength of the 34S promoter in an adapted, higher titer FMV strain would be greater in proportion to an increase in titer: this would require an increase in its strength well beyond that of the CaMV 35S promoter. The titer of virus stocks varies with host plant species, stage of the infection, environmental conditions of growth, and the strain of the virus used, while the sequence of the various promoter domains is, of course, invariant.

10. The steps in caulimoviral propagation which limit aggressiveness are not known, nor are they for other plant virus classes, but, a priori, the FMV 34S promoter does not appear to function at a step that might limit viral propagation. It is thought to be involved with the transcription of viral RNA from the supercoiled viral genomic DNA in the nucleus. This RNA goes to the cytoplasm, where it is thought to be a template for the translation of viral genes I through V. None of these gene products are prominent in infected cells. However, gene IV is the viral capsid protein. At least 60 copies of this protein are required (to build the icosohedral viral capsid) for every copy of viral replicase (gene V product) nonetheless, both of these products are thought to come from the same 34S-derived transcript. Clearly, some post-transcriptional step effects expression of these genes, and therefore the 34S-mediated transcription step is only secondarily involved with viral propagation rates.

11. The movement of the virus from initially infected cells to spread through the host is very likely important in the establishment of a high titer infection. Though the caulimoviral gene/genes involved in this are not characterized, they are most likely proteins (by analogy to systems wherein movement is understood), not promoter elements.

12. Virus capsids appear to be assembled within virus-specific cytopathic inclusions, thought to be made up, in large part, from the product of viral gene VI. Viral gene VI is likely to be the sequence which determines aggressiveness in caulimovirus infection (Daubert et al., 1983). This is the only gene in CaMV with its own dedicated promoter of RNA transcription from the supercoiled nuclear viral genome (the 19S promoter). The gene VI product is prominent, the only viral protein detectable in infected cells by standard protein-staining techniques (Young et al, 1987). The only mutations having modulating effects on viral infectivity are within gene VI (Daubert et al, 1983; Daubert & Routh, 1990). Thus, there is evidence from which we might predict that the systems involved in the expression of viral gene VI may be a predictor of viral aggressiveness. On the other hand, we have no direct evidence that changes in the FMV 34S promoter have a bearing on viral aggressiveness, or of the converse, that viral aggressiveness will be a reflection of FMV 34S promoter activity.

Executed this 13 day of May, 1993 at Davis, California.

Margaret Sanger

Margaret P. Sanger

Bibliography

Daubert et al, 1983. Insertional Mutagenesis of the CaMV genome. Gene 25,201-210

Daubert & Routh, (1990) Point mutations in CaMV gene VI. Molec.Plant Microbe Interact. 3,341-345

Guilley et al, 1982. Transcription of CaMV DNA. Cell 30,763-773.

Hasegawa et al, 1989. The complete sequence of soybean chlorotic mottle virus. N A R 17, 9993-10013

Hull et al, 1976. Purification procedure for CaMV. J.Gen.Virol. 31,93-100

Odell, et al 1985. Identification of sequences required for activity of the CaMV 35S promoter. Nature 313,810-812

Ow et al, 1987. Functional regions of the CaMV 35S promoter. PNAS 84,4870-4874

Sanger et al, (1990) Characteristics of a strong promoter from FMV. Plant Molec.Biol. 14,433-443

Sanger et al, (1991) Hypervariable domains of caulimovirus gene VI. Virology 182,830-834 Virology

Shepherd et al, 1987. FMV: properties of the virus. Phytopathol. 77,1668-1673

Young et al, 1987. Expression of CaMV proteins from an integrated form of the viral genome. J.Gen.Virol. 68,3217-3225

(19) European Patent Office
Office européen des brevets

(11) Publication number: 0 426 641 A2

EUROPEAN PATENT APPLICATION

(21) Application number: 90870202.0

(22) Date of filing: 30.10.90

(51) Int. Cl.5: C12N 15/82, C12N 5/10, C12N 15/54, C12N 15/84, A01H 5/00

(30) Priority: 31.10.89 US 429917

(43) Date of publication of application: 08.05.91 Bulletin 91/19

(84) Designated Contracting States: AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(71) Applicant: MONSANTO COMPANY
Patent Department 800 North Lindbergh Boulevard
St. Louis, Missouri 63167(US)

(72) Inventor: Rogers, Stephen Gary
52 Rue Robert Jones
B-1080 Brussels(BE)

(74) Representative: Ernst, Hubert et al
Monsanto Services International S.A., Patent Department, Avenue de Tervuren 270-272, Letter Box No. 21
B-1150 Brussels(BE)

(54) Promoter for transgenic plants.

(57) A full-length transcript promoter from figwort mosaic virus (FMV) is identified and its DNA sequence given. The promoter functions as a strong and uniform promoter for chimeric genes inserted into plant cells. This strong promoter function is exhibited by a histochemical assay in floral buds and by reproductive scores of transgenic plants including the promoter. The promoter preferably includes a 5′ leader sequence that may be from the FMV itself or from a heterologous source with respect to the promoter. The promoter is used in a plant cassette vector, a chimeric gene and in methods for transforming plant cells to obtain transgenic plants, plant cells or seeds incorporating the FMV promoter.

EP 0 426 641 A2

PROMOTER FOR TRANSGENIC PLANTS

Background of the Invention

This invention relates in general to plant genetic engineering, and more particularly, to a novel promoter for obtaining constitutive and uniform expression of chimeric genes in plants. This invention also relates to transgenic plants and plant cells containing the promoter.

One of the primary goals of plant genetic engineering is to obtain plants having improved characteristics or traits. The type and number of these characteristics or traits are innumerable, but may include virus resistance, insect resistance, herbicide resistance, enhanced stability or improved nutritional value, to name a few. Recent advances in genetic engineering have enabled researchers in the field to incorporate heterologous genes into plant cells to obtain the desired qualities in the plant of choice. This permits advantageous genes from a source different than the transformed plant to be incorporated into the plant's genome. This new gene can then be expressed in the plant cell to exhibit the new trait or characteristic.

In order for the newly inserted gene to express the protein for which it codes in the plant cell, the proper regulatory signals must be present and in the proper location with respect to the gene. These regulatory signals include a promoter region, a 5′ non-translated leader sequence and a 3′ polyadenylation sequence. The promoter is a DNA sequence that directs the cellular machinery to produce RNA. The promoter region influences the rate at which the RNA product of the gene and resultant protein product of the gene is made. The 3′-polyadenylation signal is a non-translated region that functions in plant cells to cause the addition of polyadenylate nucleotides to the 3′ end of the RNA to stabilize the RNA in the cytoplasm for subsequent translation of the RNA to produce protein.

It has previously been shown that certain promoters are able to direct RNA synthesis at a higher rate than others. These are called strong promoters. Certain promoters have also been shown to direct RNA production at higher levels only in particular types of cells and tissues. Those promoters that direct RNA production in many or all tissues of a plant are called constitutive promoters.

Previous work had shown that the 35S promoter from the cauliflower mosaic virus (CaMV35S) was the strongest constitutive promoter known in plants (Odell et al., 1985; Jensen et al., 1986; Jefferson et al., 1987; Kay et al., 1987; Sanders et al., 1987). This had been shown by demonstrating measurable levels of reporter gene proteins or mRNAs in extracts prepared from the leaves, stems, roots and flowers of transgenic plants. As a result, the CaMV35S promoter has been widely used by scientists in the field of plant genetic engineering.

Although the CaMV35S promoter appeared to be a strong, constitutive promoter in assays involving cell extracts, detailed histological analysis of a reporter gene product that is detectable at the cell and tissue level showed a rather high degree of variability of expression of this gene product. This histological analysis revealed an unknown and unexpected variability in the expression of a gene product driven by the CaMV35S promoter. This variable level and site of expression is believed to have two primary causes. The first is that variability is an intrinsic property of the CaMV35S promoter. The second is that the variability is caused by the position that the CaMV35S promoter driven DNA sequence is integrated into the genome of the transformed plant. When a gene is introduced into a plant cell, the new DNA becomes incorporated at random locations in the plant DNA. This variability in location or insert position leads to a variation in the level of promoter activity and protein production from individual transformants. As a result, a large number of individual transgenic plants must be assayed to find those that produce the highest levels of gene product in most or all of the tissues of the plants. Even the presumed strong, constitutive CaMV35S promoter is subject to the effect of insertion position variability and its use requires that a relatively large number of transformed plants be screened to find ones having appropriate levels of gene expression. Thus, it is clear that a need exists in plant genetic engineering for promoters that express high levels of chimeric gene product, but that is less subject to the wide variation in tissue level expression due to intrinsic properties of the promoter or caused by the effect of insertion position in transgenic plant DNA.

Other caulimoviruses, a group of double-stranded DNA viruses to which the cauliflower mosaic virus belongs, were considered as a potential source for such a promoter. Two caulimoviruses that are distantly related to CaMV have been previously described. The figwort mosaic virus (FMV) was described by Richins et al. (1987) and the carnation etched ring virus (CERV) was described by Hull et al. (1986). The DNA sequence and predicted gene organization of each of these two viruses were similar enough to the CaMV to permit Richins et al. to speculate as to the locations of the FMV and CERV homologues of the CaMV35S promoter. There was, however, little conservation of DNA sequences in these presumptive promoter

regions and no confirming RNA transcript analysis had been carried out to provide a demonstration of the exact location of the promoter sequences, much less a showing that a promoter from FMV would provide an increased and more uniform level of expression of a chimeric gene in plants.

It is therefore a primary object of the present invention to provide a promoter for use in transgenic plants that exhibits an increased and more uniform level of expression of a gene product driven by the promoter than that exhibited by previously known plant promoters.

It is another object of the present invention to provide a promoter for use in transgenic plants that is less affected by insertion position effects than previously known and used plant promoters.

It is a further object of the present invention to provide a promoter for use in transgenic plants that exhibits a higher level of expression of a gene product driven by the promoter in many of the tissues and cells of the plant, particularly the floral buds, than that exhibited by previously known plant promoters.

It is yet another object of the present invention to provide such a promoter for the expression of a chimeric gene in plants that is obtained from the full-length transcript of the figwort mosaic virus.

Other and further objects of the invention will be made clear or become apparent from the following description and claims when read in light of the accompanying drawings.

Brief Description of the Drawings

Figure 1 shows the DNA sequence containing the full-length transcript promoter from the figwort mosaic virus including a 5' leader sequence and a small amount of 3' flanking DNA.

Figure 2 shows a physical map of pMON721.

Figure 3 shows a physical map of pMON1573.

Figure 4 shows a physical map of pMON977.

Figure 5 shows a physical map of pMON981.

Figure 6 shows a physical map of pMON994.

Figure 7 shows the steps employed in the preparation of pMON994.

Figure 8 shows a physical map of pMON996.

Figure 9 shows the steps employed in the preparation of pMON996.

Figure 10 shows a restriction map of the T-DNA regions of the *Agrobacterium tumefaciens* strain pTiT37 plasmid which was disarmed to create the ACO *Agrobacterium* strain.

Figure 11(a) and (b) is a color photograph showing the presence of GUS activity in a tobacco flower bud transformed with the β-glucuronidase gene driven by the enhanced CaMVe35S promoter (a) and the FMV full-length transcript promoter (b).

Figure 12 shows the reproductive scores of transgenic plants containing mutant EPSPS under the control of the FMV full-length transcript promoter (PMON996) or CaMVe35S promoter (PMON899) after glyphosate application.

Summary of the Invention

It has been discovered that the full-length transcript promoter from the figwort mosaic virus (FMV) functions as a strong and uniform promoter for chimeric genes inserted into plant cells, particularly in the cells comprising the floral buds. The resulting transgenic plant expresses the protein encoded by the inserted gene at a higher and more uniform level throughout the tissues and cells of the transformed plant than the same gene driven by an enhanced CaMV35S promoter. The DNA sequence of the promoter is located between nucleotides 6368 and 6930 of the FMV genome. A 5' non-translated leader sequence is preferably coupled with the promoter. The leader sequence can be from the FMV genome itself or can be from a source other than FMV.

Other aspects of the invention include use of the FMV promoter in a method for transforming plant cells, a cassette vector including the FMV promoter, a chimeric gene including the FMV promoter sequence and transgenic plants, plant cells and seeds incorporating the FMV promoter in a chimeric gene.

Detailed Description of the Preferred Embodiment

The figwort mosaic virus (FMV) is a member of the caulimoviruses which are a group of double-stranded DNA viruses. Other members of this group include the cauliflower mosaic virus (CaMV) and the carnation etched ring virus (CERV). The CaMV and its promoter sequences are well-known in the literature (Gardner et al. 1981; Hohn et al. 1982; Guilley et al. 1982). The entire nucleotide sequence of the FMV DNA has been elucidated and reported by Richins et al. (1987). Richins et al. reported two intergenic regions in the FMV genome; a large intergenic region located between open reading frames (ORF) VI and VII and a small intergenic region located between ORFs V and VI. Richins et al. proposed that a promoter sequence analogous to the CaMV35S promoter, the major mRNA transcript promoter of the CaMV, was located in the large intergenic region of the FMV genome, but no confirming RNA transcript analysis had been carried out to provide a demonstration of the exact location of the transcriptional start and, consequently, the promoter sequence.

One aspect of the present invention includes isolation of the promoter for the full-length transcript from the figwort mosaic virus and the determination of the sequence of this promoter. The promoter preferably includes a 5' leader sequence that may be from the FMV promoter sequence itself or from a source heterologous with respect to the promoter.

The novel promoter of the instant invention was isolated from a small DNA fragment from a complete, full-length clone of FMV DNA. A plasmid, pFMVSc3, was obtained from Dr. R.J. Shepherd of the University of Kentucky. The nucleotide sequence of the FMV DNA and the organization of the FMV genome are given in Richins et al. (1987). This plasmid contains the complete DNA from FMV as adapted for growth on solanaceous hosts as described in Shepherd et al. (1987). As a result of the adaptation of the FMV DNA for growth on solanaceous hosts, the FMV DNA is believed to have undergone a number of mutations at the nucleotide level. In the description and examples that follow FMV DNA from such an adapted strain is used. It is to be understood that the teachings and examples of this invention would also apply to a promoter region isolated from a "wild-type" or non-adapted FMV DNA with similar advantages and results. The original virus was isolated from *Scrophularia californica*. The FMV DNA was cloned into the unique SacI site of pUC13 (Vieira, J. and Messing, J., 1982) to obtain pFMVSc3. The nucleotide sequences shown in the drawing figures accompanying this disclosure that relate to FMV follow the numbering system used by Richins et al.

The FMV promoter sequence was isolated by digesting pFMVSc3 with sspI which cleaves the FMV DNA at several sites including between nucleotides 6367 and 6368 and between nucleotides 6948 and 6949. This releases a 581 base pair (bp) nucleotide fragment that contains a promoter sequence and 18 nucleotides of 5' non-translated leader sequence corresponding to the full-length transcript promoter of FMV. The nucleotide sequence of this fragment and a small amount of flanking DNA is shown in Fig. 1.

This fragment was purified using the NA-45 membrane method after electrophoretic separation on a 0.8% agarose gel and inserted into plasmid pMON721 that had been cleaved with StuI. A physical map of pMON721 is shown in Fig. 2.

As shown in Fig. 2, plasmid pMON721 contains a StuI site in a multilinker flanked by a HindIII site on one side and a BglII site on the other side. Once the SspI fragment was inserted into pMON721 at the StuI site, the resulting transformed pMON721 plasmids were screened for identification of transformants carrying the presumed FMV full-length RNA transcript promoter fragment oriented in the proper manner. A plasmid identified as pMON1573 was identified as containing the FMV promoter fragment properly oriented so that the presumed 5' or upstream sequences of the promoter were adjacent to the HindIII site and the untranslated leader sequences terminated at the BglII site. Fig. 3 is a physical map of pMON1573.

Once a plasmid containing the FMV major RNA (full-length) transcript promoter sequence in the correct orientation was isolated, a cassette vector containing this promoter was prepared. A cassette vector is a cloning vector that typically includes all of the necessary elements needed for transformation of plants or plant cells. Typical plant cloning vectors comprise selectable and scoreable marker genes, T-DNA borders, cloning sites, appropriate bacterial genes to facilitate identification of transconjugates, broad host-range replication and mobilization functions and other elements as desired. A cassette vector containing the FMV major RNA transcript promoter of the present invention in a suitable plant transformation vector was prepared by starting with the pMON977 plasmid. A physical map of pMON977 is as illustrated in Fig. 4.

As shown in Fig. 4, pMON977 has the following elements; a 0.93 kb fragment isolated from transposon Tn7 encoding a bacterial spectinomycin streptomycin resistance gene (Spc/Str) that functions as a marker for selection of the plasmid in *E. coli* and *Agrobacterium* (Fling, M.E., et al. 1985); a 1.61 kb segment of DNA encoding a chimeric kanamycin resistance gene (P-35S Kan NOS3') that permits selection of transformed plant cells (Beck, E., et al. 1982); a 0.75 kb oriV DNA sequence containing the origin of replication from the Rk2 plasmid (Stalker, D.M., et al. 1979) a 3.1 kb segment of pBR322 (ori-322) that provides the origin of replication for maintenance in *E. coli* and the bom site for the conjugational transformation to the *Agrobacterium* cells (Sutliffe, J., 1979); a 0.36 kb segment from pTiT37 (the PvuI to BclI fragment) that carries the nopaline-type T-DNA right border (Fraley et al. 1985); and a 1.15 kb expression cassette consisting of the 0.66 kb enhanced 35S promoter P-e35S (Kay et al. 1987), several unique restriction sites and the 0.7 kb 3' non-translated region of the pea ribulose bisphosphate carboxylase small subunit E9 gene (E9 3') (Coruzzi, G., et al., 1984 and Morelli, G. et al., 1985). Plasmid pMON977 was cut with HindIII and BglII to remove the CaMV P-e35S enhanced 35S promoter. A 605bp fragment containing the FMV full-length transcript promoter was excised from pMON1573 with HindIII and BglII and cloned into pMON977 to create pMON981. Plasmid pMON981 thus contains the FMV full-length transcript promoter and the E9-3' gene (FMV-E9 3') as an expression cassette. Also included in pMON981

between the FMV promoter and the E9-3' gene are restriction endonuclous sites for XbaI, BglII and SmaI. A physical map of pMON981 is shown in Fig. 5.

In order to determine that the isolated FMV sequence included the desired promoter region and to demonstrate the effectiveness and utility of the isolated FMV promoter, reporter genes were inserted into plant cassette vector pMON981. The reporter genes chosen were the *E. coli* β-glucuronidase (GUS) coding sequence and the *Arabidopsis* EPSP synthase gene containing a single glycine to alanine substitution which causes this enzyme to be tolerant of glyphosate herbicides.

The *E. coli* β-glucuronidase coding sequence was inserted into the unique BglII site in the FMV-E9 3' cassette of plasmid pMON981. The GUS gene was excised from pMON637 on an 1885bp BglII to BamHI fragment. The resulting plasmid was denoted pMON994 and contains the GUS gene under control of the FMV promoter. Plasmid pMON994 is shown in Fig. 6 and a flow chart illustrating the development of pMON994 is shown in Fig. 7.

EPSP synthase (5-enolpyruvyl-3-phosphoshikimate synthase; EC:25.1.19) is an enzyme involved in the shikimic acid pathway of plants. The shikimic acid pathway provides a precursor for the synthesis of aromatic amino acids essential to the plant. Specifically, EPSP synthase catalyzes the conversion of phosphoenol pyruvate and 3-phosphoshikimic acid to 5-enolpyruvyl-3-phosphoshikimic acid. A herbicide containing N-phosphonomethylglycine inhibits the EPSP synthase enzyme and thereby inhibits the shikimic acid pathway of the plant. The term "glyphosate" is usually used to refer to the N-phosphonomethylglycine herbicide in its acidic or anionic forms. Novel EPSP synthase enzymes have been discovered that exhibit an increased tolerance to glyphosate containing herbicides. In particular, an EPSP synthase enzyme having a single glycine to alanine substitution in the highly conserved region having the sequence: -L-G-N-A-G -T-A- located between positions 80 and 120 in the mature wild-type EPSP synthase amino acid sequence has been shown to exhibit an increased tolerance to glyphosate and is described in the commonly assigned pending patent application entitled "Glyphosate-Tolerant 5-Enolpyruvyl-3-Phosphoshikimate Synthase" having U.S. serial number 931,492, the teachings of which are hereby incorporated by reference hereto. Methods for transforming plants to exhibit glyphosate tolerance are discussed in the commonly assigned U.S. patent application entitled "Glyphosate-Resistant Plants," Serial No. 879,814 filed July 7, 1986, the disclosure of which is specifically incorporated synthase plant gene encodes a polypeptide which contains a chloroplast transit peptide (CTP) which enables the EPSP synthase polypeptide (or an active portion thereto) to be transported into a chloroplast inside the plant cell. The EPSP synthase gene is transcribed into mRNA in the nucleus and the mRNA is translated into a precursor polypeptide (CTP-mature EPSP synthase) in the cytoplasm. The precursor polypeptide is transported into the chloroplast.

The EPSP synthase gene containing a single glycine to alanine mutation obtained from mutated *Arabidopsis thaliana* gene sequence was also inserted into the FMV-E9 3' cassette vector of plasmid pMON981. Plasmid pMON981 was cut with XbaI and SmaI. The *Arabidopsis* EPSP synthase gene is located on plasmid pMON897. Plasmid pMON897 is obtained by excising the *Arabidopsis* EPSP synthase gene (AEPSPS) in pMON600 by cutting with ClaI and EcoRI. This fragment is inserted into pMON855 which includes a multilinker containing sites for EcoRI, ClaI and XbaI. Plasmid pMON855 is cut with ClaI and EcoRI and the *Arabidopsis* EPSP synthase fragment isolated from pMON600 is inserted. The resulting plasmid is pMON897. Plasmid pMON897 was then cut with EcoRI and the ends were filled in using Klenow polymerase and then cut with XbaI and the *Arabidopsis* EPSP synthase gene was excised as a 3881 bp fragment. The *Arabidopsis* EPSP synthase gene was then cloned into pMON981 digested with XbaI and SmaI to create pMON996. A physical map of pMON996 is shown in Fig. 8 and a flow chart illustrating the development of pMON996 is shown in Fig. 9.

Once the FMV-E9 3' cassette vector containing the desired reporter gene is prepared, the vector can then be inserted into suitable *Agrobacterium* strains for *Agrobacterium* mediated transformation into plants or plant cells. The *Agrobacterium tumefaciens* strain to be used preferably contains a disarmed Ti plasmid. Two particularly useful strains are *Agrobacterium tumefaciens* strain A208 carrying the disarmed Ti plasmid pTiC58 derivative, pMP90RK (Koncz and Schell, 1986) and the ACO *Agrobacterium tumefaciens* strain carrying the disarmed pTiT37-CO nopaline type plasmid.

The *A. tumefaciens* strain 208 carrying the disarmed pMP90RK plasmid does not carry the T-DNA phytohormone genes and therefore cannot cause crown gall disease. When this strain is used for plant transformations, the vector plasmid is introduced into the *Agrobacterium* by the triparental conjugation system (Ditta et al. 1980) using the helper plasmid pRK2013. The vectors are transferred to plant cells by the <u>vir</u> functions encoded

by the disarmed pMP90RK Ti plasmid. Analysis of transformants suggest that the vector is opened at the pTiT37 right border sequence and the entire vector sequence is inserted into the host plant chromosome. The pMP90RK Ti plasmid is probably not transferred to the plant cell but remains in the *Agrobacterium*.

Figure 10 shows a restriction map of the T-DNA regions of the *Agrobacterium tumefaciens* strain pTiT37 plasmid which was disarmed to create the ACO *Agrobacterium* strain. This strain carries the disarmed pTiT37-CO nopaline type plasmid. The hatched boxes in Fig. 10 show the segments of the Ti plasmid DNA which were used to provide homology for recombination and replacement of the T-DNA. The T-DNA segment was replaced by the Tn601 bacteria kanamycin resistance gene (Kn^R) segment joined to the OriV and pBR322 segment homologous to the vectors described above. The recombination between the disarmed pTiT37-CO and plant cassette vector takes place through the pBR322 oriV area of homology resulting in the hybrid T-DNA which contains the entire DNA of the cassette vector plasmid. On cultivation of the *Agrobacterium* with plant cells, the hybrid T-DNA segment between the left and right borders is transferred to the cells and integrated into the genomic DNA.

Once the vector has been introduced into the disarmed *Agrobacterium* strain, the desired plant can then be transformed. Any known method of transformation that will work with the desired plant can be utilized. These methods include the leaf disc method of Horsch et al. (1984) and as adapted by Fry et al. (1986) for *Brassica napus*. Also conceived to be within the scope of the present invention is the use of DNA fragments or vectors including the FMV promoter sequences coupled with heterologous DNA sequences in the transformation of plants utilizing techniques such as electroporation or particle gun transformation.

Suitable plants for the practice of the present invention include, but are not limited to, soybean, cotton, alfalfa, oilseed rape, flax, tomato, sugar beet, sunflower, potato, tobacco, maize, wheat, rice and lettuce.

The effectiveness of the FMV promoter was determined by comparison studies with the enhanced CaMV35S promoter. In one study, pMON994 containing the FMV promoter including the 5′ non-translated leader sequence from FMV fused to the β-glucuronidase reporter gene and the E9-3′ non-translated polyadenylation region from pea was introduced into tobacco using the leaf disc method of Horsch et al. (1984) and transgenic plants obtained.

Tobacco (*Nicotiani tabacum var. samsun*) leaf disks with diameters of about 6mm (¼inch) were taken from surface sterilized tobacco leaves. Those were cultivated on MS104 agar medium for two days to promote partial cell wall formation at the wound surfaces. They were then submerged in a culture of *A. tumefaciens* cells containing both pMON994 and pMP90RK which had been grown overnight in Luria broth at 28°C. and shaken gently. The cells were removed from the bacterial suspension, blotted dry, and incubated upside down on filter paper placed over "nurse" cultures of tobacco cells as described by Horsch (1980). After two or three days, the disks were transferred to petri dishes containing MS media with 500μg/ml carbenicillin with no nurse culture.

Control tissue was created using *A. tumefaciens* cells containing the helper plasmid pMP90RK and a different plant transformation vector, pMON505, which contained a T-DNA region with a NOS/NPTII/NOS kanamycin resistance gene and a NOS selectable marker gene identical to pMON994, but without the FMV/β-glucuronidase gene.

Within ten days after transfer to the MS media, actively growing callus tissue appeared on the periphery of all disks on both the control and transformed plates.

Transformed tobacco plants were produced by regeneration from the above-described transformed leaf disks by the procedure described by Horsch, et al. (1985). The transformed plants obtained contained the pMON994 vector which contains the FMV promoter fused to the β-glucuronidase gene.

The same procedure as described above was utilized to obtain transformed tobacco plants containing the enhanced CaMV35S (CaMVe35S or P-e35S) promoter fused to the β-glucuronidase reporter gene and the E9-3′ non-translated polyadenylation region from pea.

A second study involved obtaining transformed canola plants (Brassica napus) carrying the *Arabidopsis* EPSP synthase gene containing a single glycine to alanine substitution at amino acid 101 driven by either the FMV promoter or the CaMVe35S promoter. The pMON996 plasmid carrying the *Arabidopsis* EPSP synthase gene directed by the FMV promoter was introduced into canola by the method of Fry et al. (1986). Four terminal internodes from plants just prior to bolting or in the process of bolting, but before flowering were removed and surface sterilized in 70° v/v ethanol for one minute, 2% w/v sodium hypochlorite for twenty minutes, and rinsed three times in sterile distilled water. Stem segments were cut into 5mm discs (Stringam 1977) and placed in a sterile 15x100mm petri plate, noting the orientation of the basal end. The discs were inoculated for five minutes by pouring two to four milliliters of an overnight culture of the ACO *A. tumefaciens* strain

containing pMON996 as previously described over the discs in the petri plate and then blotted dry by placing sterile filter paper in the petri plate and turning the plate over to absorb any excess bacteria. The stem discs were placed basal side down on feeder plates on medium containing 1/10x standard MS salts, B5 vitamins, 3% sucrose, 0.8% agar, pH 5.7, 1mg/l BA and 1.4ml TXD feeder cells (Horsch et al. 1985).

After a two to three day coculture period, stem discs were transferred, five to a deep dish petri plate (25 x 100mm) containing the same medium with standard MS salts, 1mg/l BA, 500 mg/l carbenicillin, 0.3 mm arginine, and 100 mg/l kanamycin for selection. At three weeks the stem explants were transferred to fresh plates containing the same medium. Culture of the explants was in a growth room under continuous cool white light at 26°C. Shoots that developed in the next one to three week period were excised from the stem explants, dipped in Rootone® and placed in 2½ inch pots containing water saturated Metro Mix 350 in closed GAF containers for ten days in a chamber with a constant temperature of 21°C and a 16 hour photoperiod. The shoots are assayed for the presence of kanamycin resistance immediately after being excised from the stem explant while still sterile.

This same procedure was used to obtain transformed canola plants containing the enhanced CaMVe35S promoter fused to the *Arabidopsis* EPSP synthase gene by inoculating the stem segment discs with ACO *Agrobacterium tumefaciens* strain containing pMON899.

Example 1

Transformed plants containing the GUS gene driven by either the FMV full-length promoter or the enhanced CaMVe35S promoter were assayed using a histological staining procedure to determine GUS activity in the transformed cells. The results of these assays on plants transformed with pMON994 (FMV/GUS/E9) were compared to the results of the same assays performed on plants transformed with pMON977 (CaMVe35S/GUS/E9).

The histochemical assay of the tobacco plants containing the FMV/GUS/E9 and CaMVe35S/GUS/E9 constructs involved examination of young flower bud (10mm) sections of the transformed plants to determine GUS activity. The flower bud section of the transformed plant was prepared by using a razor blade to free-hand section the plant tissue into sections less than 0.5mm in thickness. The tissue was then placed in excess X-gluc solution so that the section was fully covered. Pulling a vacuum on the sections may aid in penetration of the X-gluc solution. A 50ml X-gluc solution was prepared by combining 25ml of 0.2M $NaPO_4$ buffer pH 7.0, 24.0ml dH_2O, 0.25ml 0.1M $K_3[Fe(CN)_6]$, 0.25ml 0.1M $K_4[Fe(CN)_6]$ and 0.5ml 1M EDTA,pH7.0. To this solution, 50mg of X-gluc (5-bromo-4-chloro-3-idolyl-β-glucuronide) obtained from Research Organics (Cleveland, Ohio) was added and stirred until dissolved. The solution was then preferably sterilized by filtration. The flower bud sections in the X-gluc solution were then placed at 37°C for 2-4 hours. Care was taken to prevent evaporation of the solution. After the incubation period, the sections were rinsed with phosphate buffer, or distilled H_2O, and the sections were examined immediately with a dissecting scope or compound microscope. If there is interference from the pigments, the tissue can be fixed in FAA solution (85ml 50% ethanol, 5ml glacial acetic acid and 10ml formalin) for 24 hours. Problems with phenolics can be mitigated by the addition of sodium metabisulfite to 20mM to the staining solution just prior to staining. Figure 11 illustrates the results of the histological staining assay of the FMV containing GUS construct and the CaMVe35S containing GUS construct, respectively.

A positive test for the presence of GUS activity is shown by a blue coloration appearing in the tissue of the assayed plant section. In Fig. 11, a color photograph of the stained section of a tobacco flower bud transformed with the β-glucuronidase gene driven by the enhanced CaMVe35S promoter (a) and the FMV full-length promoter (b) is shown. Fig. 11(a) exhibits a typical staining profile for a CaMVe35S promoter driven GUS gene with staining in some tissues and no staining in other tissues within a single transgenic plant. The level of expression in those tissues expressing the GUS gene is considered fair. In Fig. 11(b), tissue from a plant transformed with the FMV promoter driven GUS gene shows that the transformed plant is showing much higher levels of GUS expression and a more uniform pattern of expression throughout the tissue and cells. This is illustrated by the predominant blue coloration throughout the section.

The distribution of expression and the number of highly expressing transgenic plants obtained show that the FMV promoter is superior in tissue distribution and uniformity of expression when compared to the best enhanced CaMV promoter containing transformed plants. Greater than 90% of the FMV/GUS containing transformed plants showed very strong GUS expression and that the staining was uniform from plant to plant and tissue to tissue. This staining is consistently as good in the FMV containing plants as that in the best enhanced CaMV GUS plants identified.

Example 2

Transgenic plants containing the *Arabidopsis* EPSP synthase gene containing a single glycine to alamine mutation at nucleotide 101 driven by either the FMV promoter or the CaMVe35S promoter were obtained and analyzed for resistance to glyphosate. The transgenic plants containing the *Arabidopsis* EPSP synthase gene (as described) directed by the FMV promoter contained pMON996 while those plants containing the enhanced CaMVe35S promoter contained pMON899. These transgenic plants were planted and the seed from the R_0 plants harvested, threshed and dried before planting for a glyphosate spray test. The progeny were planted in 4-inch square pots of Metro 350 and three types of slow release fertilizers. A goal of twenty seedlings from each R_0 plant is desirable for testing. Germination frequency is usually high but overplanting ensures that twenty seedlings are present. The plants were thinned down by selecting the twenty most vigorous and erect seedlings seven to ten days after planting. A negative control (non-transformed, "Westar" variety) was planted at the same time to maintain quality and display the results. The plants were maintained and grown in a greenhouse environment. A sixteen-hour photoperiod and a temperature of 21°C (day) and 15°C (night) was maintained. Water soluble Peters Pete Lite fertilizer with an analysis of 20-19-18 was applied once per week or as needed.

Two plants from each R_0 progeny were not sprayed and served as controls to compare and measure the glyphosate tolerance. When the remaining plants reached the six to eight leaf stage, usually 20 to 28 days after planting, glyphosate was applied at a rate equivalent to 0.28 Kg/ha. Low rate technology using low volumes has been adopted. A volume of ten imperial gallons for 0.28 Kg/ha of glyphosate is standard in field tests. A laboratory test sprayer had been calibrated to deliver a consistent rate equivalent to field conditions.

Results of reproductive evaluations are shown in Fig. 12. These calculations are based upon a numerical scoring system relative to nonsprayed controls. Reproductive scores are examined at 28 days after spraying and are based upon six distinct conditions in which the main meristem or flowers reacted to the glyphosate. The scale used is:

0 = no floral bud development
2 = floral buds, but aborted prior to opening
4 = flowers without antlers, antlers should protrude past petals
6 = flowers with normal appearing antlers, but sterile
8 = flowers with partially sterile antlers
10 = fully fertile flowers

Figure 12 compares the reproductive scores of the total number of transgenic canola lines containing the FMV promoter with transgenic lines containing the CaMVe35S promoter. As can be seen in Fig. 12, the reproductive scores of three of the seven transgenic lines containing the FMV promoter (pMON996) are better than any of the scores from lines containing the CaMVe35S promoter (pMON899). In fact, the transgenic lines containing pMON899 used in Fig. 12 exhibit the highest levels of glyphosate tolerance among 150 lines previously tested. This demonstrates that the FMV promoter more uniformly expresses a gene product throughout the tissues and cells of the plant, and particularly in the floral buds. It is to be understood that an increased level of expression in the floral buds is important for maximal glyphosate resistance.

The embodiments and examples described above are provided to better elucidate the practice of the present invention. It should be understood that these embodiments and examples are provided for illustrative purposes only, and are not intended to limit the scope of the invention.

References

Beck, E., Ludwig, G., Auerswald, E.A., Reiss, B., Shaller, H., Gene, 19:327 (1982)

Coruzzi, G., Broglie, R., Edwards, C., Chua, N.H., EMBO J., 3:1671 (1984)

Fling, M.E., Kopf, J., Richards, C., NAR, 13:7095 (1985)

Fry, J., Barnason, A. and Horsch, R., Plant Cell Reports, 6:321-325 (1987)

Gardner, R.C. et al., Nucleic Acids Research, Vol. 9, No. 12:287 (1981)

Guilley, H. et al., Cell, 30:763 (1982)

Hohn, T. et al., in Gene Cloning in Organisms Other than E. coli, p.193, Hofschneider and Goebel, eds. (Springer Verlag, N.Y., 1982)

Horsch R. and Jones G., In Vitro, 16:103-108 (1980)

Horsch R., Fry J., Hoffman, N., Wallworth, M., Eicholtz, D., Rogers, S., Fraley, R., Science 227:1229-1231 (1985)

Hull, R., Sadler, J., Longstaff, M., EMBO Journal, Vol. 5, No. 12:3083-3090 (1986)

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W., EMBO J., 6:3901-3907 (1987)

Jensen, J.S., Marcker, K.A., Otten, L. and Schell, J., Nature, 321:669-674 (1986)

Kay, R., Chan, A., Daly, M., McPherson, J., Science, 236:1299-1302 (1987)

Morelli, G., Nagy, F., Fraley, R.T., Rogers, S.G., Chua, N.H., Nature Vol. 315, pp.200-204 (1985)

Odell, J.T., Nagy, F. and Chua, N.H., Nature,

313:810-812 (1985).
Richins, R., Scholthof, H., Shepherd, R., Nucleic Acid Research , 15, No. 20:8451-8466 (1987)
Sanders, P.R., Winter, J.A., Barnason, A.R., Rogers, S.G. and Fraley, R.T., Nucleic Acids Research , 4:1543-1558 (1987)
Shepherd, R., Richins, R., Duffus, J., Hadloy, M., Phytopathology , Vol. 77, No. 12:1668-1673 (1987)
Stalker, D.M., Kolter, R., Helinski, D., Mol. Gen. Genet. , 181:8 (1979)
Stringam, G.R., Plant Science Letters , 9:115-119 (1977)
Sutliffe, J., Cold Spring Harbor Symposium, 43:77 (1979)
Vieira, J. and Messing, J., Gene 19:259 (1982); Biochemicals for Molecular Biology , p.126 (1989)

Claims

1. A full-length transcript promoter from figwort mosaic virus.
2. The promoter of claim 1 wherein said promoter has the nucleotide sequence as shown in nucleotides 6368 to 6930 of Figure 1.
3. A promoter of claim 1 further comprising a 5′ non-translated leader sequence from figwort mosaic virus.
4. A promoter of claim 1 further comprising a 5′ non-translated leader sequence from a source heterologous with respect to the promoter.
5. The promoter of claim 3 wherein said promoter has the nucleotide sequence as shown in Figure 1.
6. A method for transforming a plant cell to express a chimeric gene, the improvement comprising a chimeric gene containing a full-length transcript promoter from figwort mosaic virus.
7. A method of claim 6 wherein said promoter includes a 5′ non-translated leader sequence.
8. A method of claim 7 wherein said 5′ non-translated leader sequence is from figwort mosaic virus.
9. A method of claim 7 wherein said 5′ non-translated leader sequence is from a source heterologous with respect to the promoter.
10. A chimeric gene that functions in plant cells comprising:
a full-length transcript promoter from figwort mosaic virus;
a structural DNA sequence that is heterologous with respect to the promoter; and
a 3′ non-translated region which encodes a polyadenylation signal which functions in plant cells to cause the addition of polyadenylate nucleotides to the 3′ end of the RNA.
11. The chimeric gene of claim 10 wherein said promoter further comprises a 5′ non-translated leader sequence.
12. The chimeric gene of claim 11 wherein said 5′ non-translated leader sequence is from figwort mosaic virus.
13. The chimeric gene of claim 11 wherein said 5′ non-translated leader sequence is from a source heterologous with respect to the promoter.
14. A chimeric gene of claim 10 wherein said structural DNA sequence comprises a coding sequence which causes the production of RNA, encoding a chloroplast transit peptide/5-enolpyruvylshikimate-3-phosphate synthase fusion polypeptide, which chloroplast transit peptide permits the fusion polypeptide to be imported into a chloroplast of a plant cell.
15. A transformed plant cell that contains a chimeric gene comprising:
a full-length transcript promoter from figwort mosaic virus;
a structural DNA sequence that is heterologous with respect to said promoter; and
a 3′ non-translated region which encodes a polyadenylation signal which functions in plant cells to cause the addition of polyadenylate nucleotides to the 3′ end of the RNA.
16. A plant cell of claim 15 wherein said promoter further comprises a 5′ non-translated leader sequence.
17. A transformed plant cell of claim 16 wherein said 5′ non-translated leader sequence is from figwort mosaic virus.
18. A transformed plant cell of claim 16 wherein said 5′ non-translated leader sequence is from a source heterologous with respect to the promoter.
19. A transformed plant cell of claim 15 wherein said structural DNA sequence further comprises a coding sequence which causes the production of RNA, encoding a chloroplast transit peptide/5-enolpyruvylshikimate-3-phosphate synthase fusion polypeptide, which chloroplast transit peptide permits the fusion polypeptide to be imported into a chloroplast of a plant cell.
20. A transformed plant cell of claim 19 wherein the coding sequence encodes a glyphosate tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS).
21. A transformed plant cell of claim 19 wherein the chloroplast transit peptide is from a plant EPSPS gene.
22. A chimeric gene of claim 14 wherein the coding sequence encodes a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS).
23. A chimeric gene of claim 14 wherein the chloroplast transit peptide is from a plant EPSPS gene.
24. A plant transformation vector which comprises

a disarmed plant tumor inducing plasmid of Agrobacterium tumefaciens which is capable of inserting a chimeric gene into susceptible plant cells, wherein said chimeric gene comprises a full-length transcript promoter from figwort mosaic virus and a structural DNA sequence that is heterologous with respect to the promoter.

25. A plant transformation vector of claim 24 wherein said promoter further comprises a 5′ non-translated leader sequence.

26. A plant transformation vector of claim 25 wherein said 5′ non-translated leader sequence is from figwort mosaic virus.

27. A plant transformation vector of claim 25 wherein said 5′ non-translated leader sequence is from a source heterologous with respect to the promoter.

28. A transgenic plant which comprises plant cells of any of Claims 15 to 21.

29. A seed from a plant of Claim 28.

FIG. 1

```
                 S
                 s
                 p
                 I
6358  TCATCAAAATATTTAGCAGCATTCCAGATTGGGTTCAATCAACAAGGTACGAGCCATATC
      ---------+---------+---------+---------+---------+---------+  6417
      AGTAGTTTTATAAATCGTCGTAAGGTCTAACCCAAGTTAGTTGTTCCATGCTCGGTATAG

6418  ACTTTATTCAAATTGGTATCGCCAAAACCAAGAAGGAACTCCCATCCTCAAAGGTTTGTA
      ---------+---------+---------+---------+---------+---------+  6477
      TGAAATAAGTTTAACCATAGCGGTTTTGGTTCTTCCTTGAGGGTAGGAGTTTCCAAACAT

6478  AGGAAGAATTCTCAGTCCAAAGCCTCAACAAGGTCAGGGTACAGAGTCTCCAAACCATTA
      ---------+---------+---------+---------+---------+---------+  6537
      TCCTTCTTAAGAGTCAGGTTTCGGAGTTGTTCCAGTCCCATGTCTCAGAGGTTTGGTAAT

6538  GCCAAAAGCTACAGGAGATCAATGAAGAATCTTCAATCAAAGTAAACTACTGTTCCAGCA
      ---------+---------+---------+---------+---------+---------+  6597
      CGGTTTTCGATGTCCTCTAGTTACTTCTTAGAAGTTAGTTTCATTTGATGACAAGGTCGT

6598  CATGCATCATGGTCAGTAAGTTTCAGAAAAAGACATCCACCGAAGACTTAAAGTTAGTGG
      ---------+---------+---------+---------+---------+---------+  6657
      GTACGTAGTACCAGTCATTCAAAGTCTTTTTCTGTAGGTGGCTTCTGAATTTCAATCACC

6658  GCATCTTTGAAAGTAATCTTGTCAACATCGAGCAGCTGGCTTGTGGGGACCAGACAAAAA
      ---------+---------+---------+---------+---------+---------+  6717
      CGTAGAAACTTTCATTAGAACAGTTGTAGCTCGTCGACCGAACACCCCTGGTCTGTTTTT

6718  AGGAATGGTGCAGAATTGTTAGGCGCACCTACCAAAAGCATCTTTGCCTTTATTGCAAAG
      ---------+---------+---------+---------+---------+---------+  6777
      TCCTTACCACGTCTTAACAATCCGCGTGGATGGTTTTCGTAGAAACGGAAATAACGTTTC

6778  ATAAAGCAGATTCCTCTAGTACAAGTGGGGAACAAAATAACGTGGAAAAGAGCTGTCCTG
      ---------+---------+---------+---------+---------+---------+  6837
      TATTTCGTCTAAGGAGATCATGTTCACCCCTTGTTTTATTGCACCTTTTCTCGACAGGAC

6838  ACAGCCCACTCACTAATGCGTATGACGAACGCAGTGACGACCACAAAAGAATTCCCTCTA
      ---------+---------+---------+---------+---------+---------+  6897
      TGTCGGGTGAGTGATTACGCATACTGCTTGCGTCACTGCTGGTGTTTTCTTAAGGGAGAT

                                                   S
                                                   s
                                                   p
                                                   I
6898  TATAAGAAGGCATTCATTCCCATTTGAAGGATCATCAGATACTGAACCAATATTTCTC
      ---------+---------+---------+---------+---------+--------+  6955
      ATATTCTTCCGTAAGTAAGGGTAAACTTCCTAGTAGTCTATGACTTGGTTATAAAGAG
```

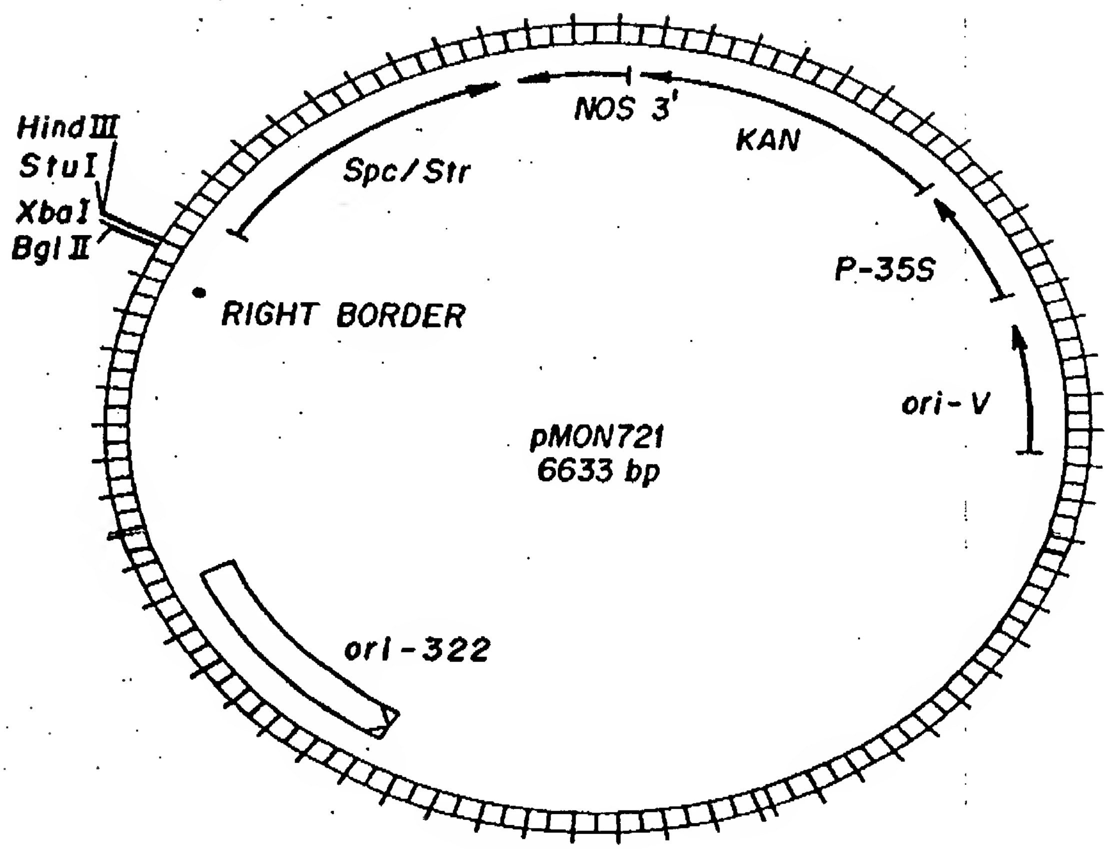

HindIII
StuI
XbaI
BglII
Spc/Str
NOS 3'
KAN
P-35S
RIGHT BORDER
ori-V
pMON721
6633 bp
ori-322

FIG. 2

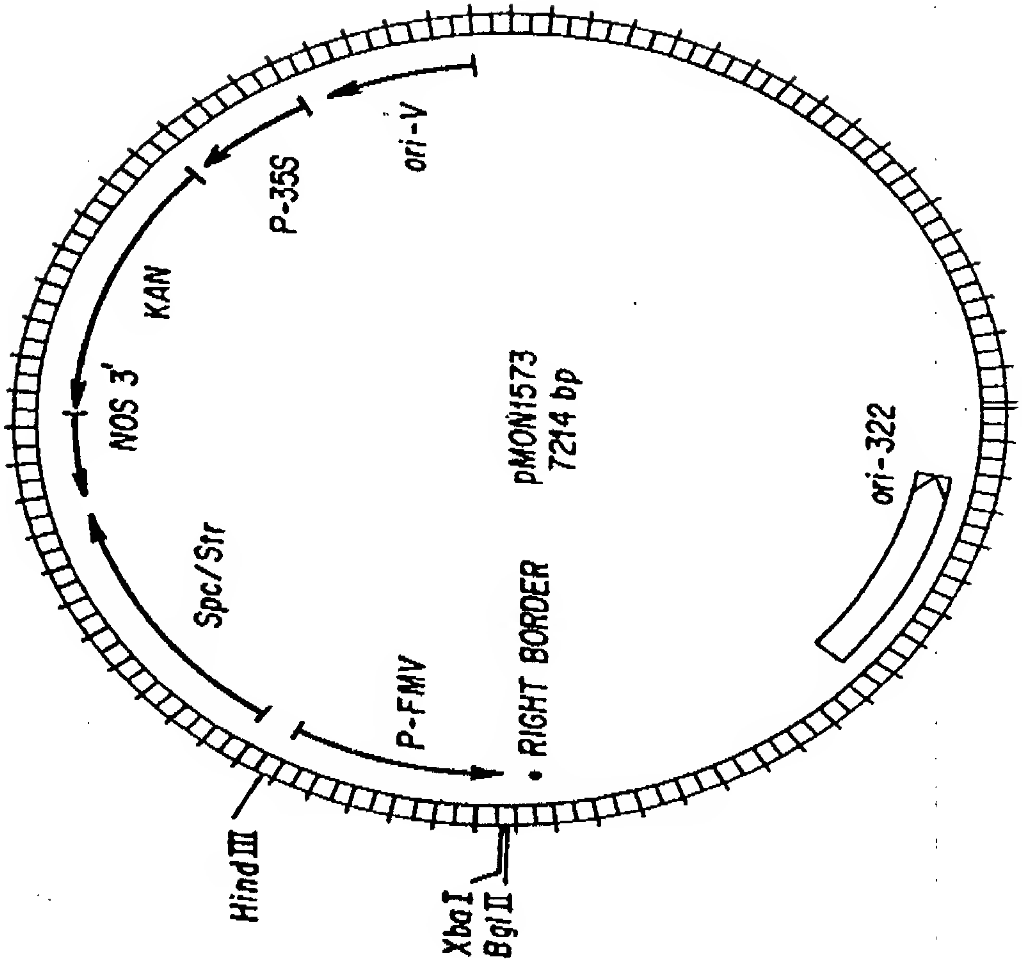

ori-V
P-35S
KAN
NOS 3'
Spc/Str
P-FMV
RIGHT BORDER
pMON1573
7214 bp
ori-322
HindIII
XbaI
BglII

FIG. 3

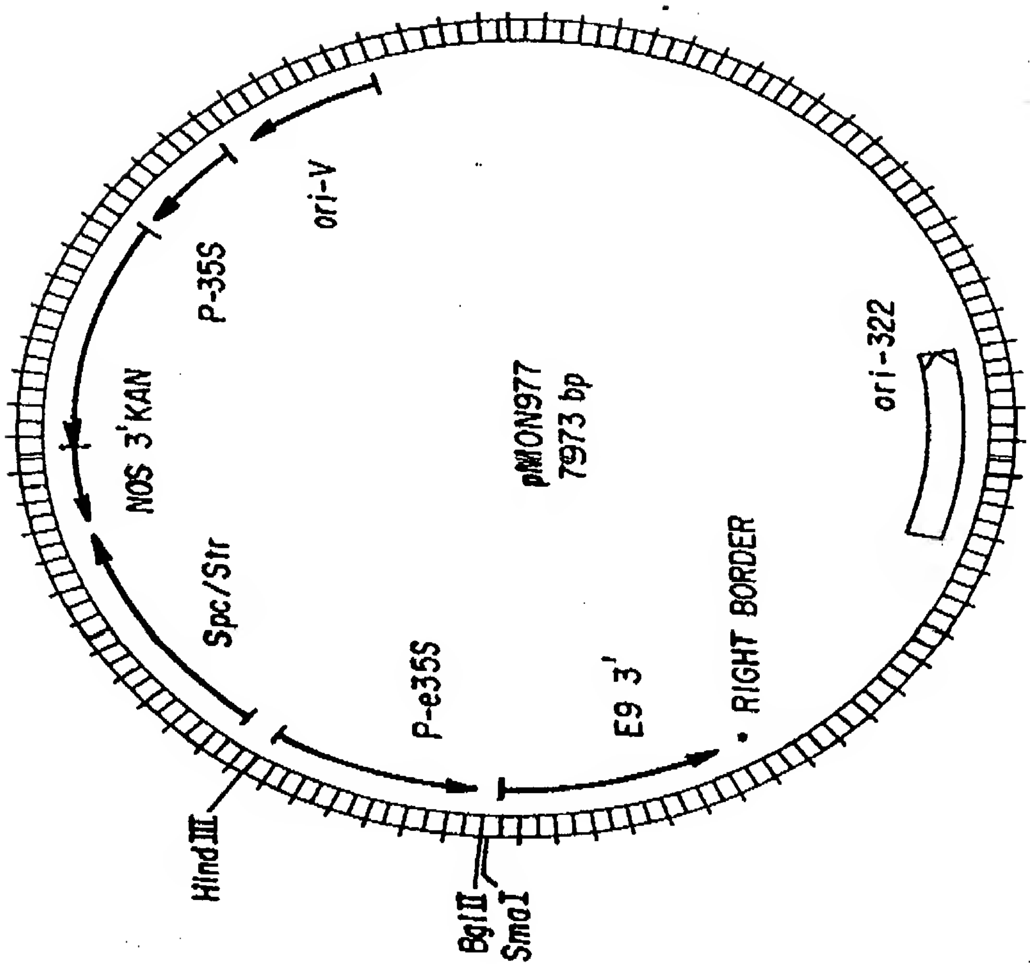

ori-V
P-35S
NOS 3'KAN
Spc/Str
P-e35S
E9 3'
RIGHT BORDER
ori-322
pMON977
7973 bp
HindIII
BglII
SmaI

FIG. 4

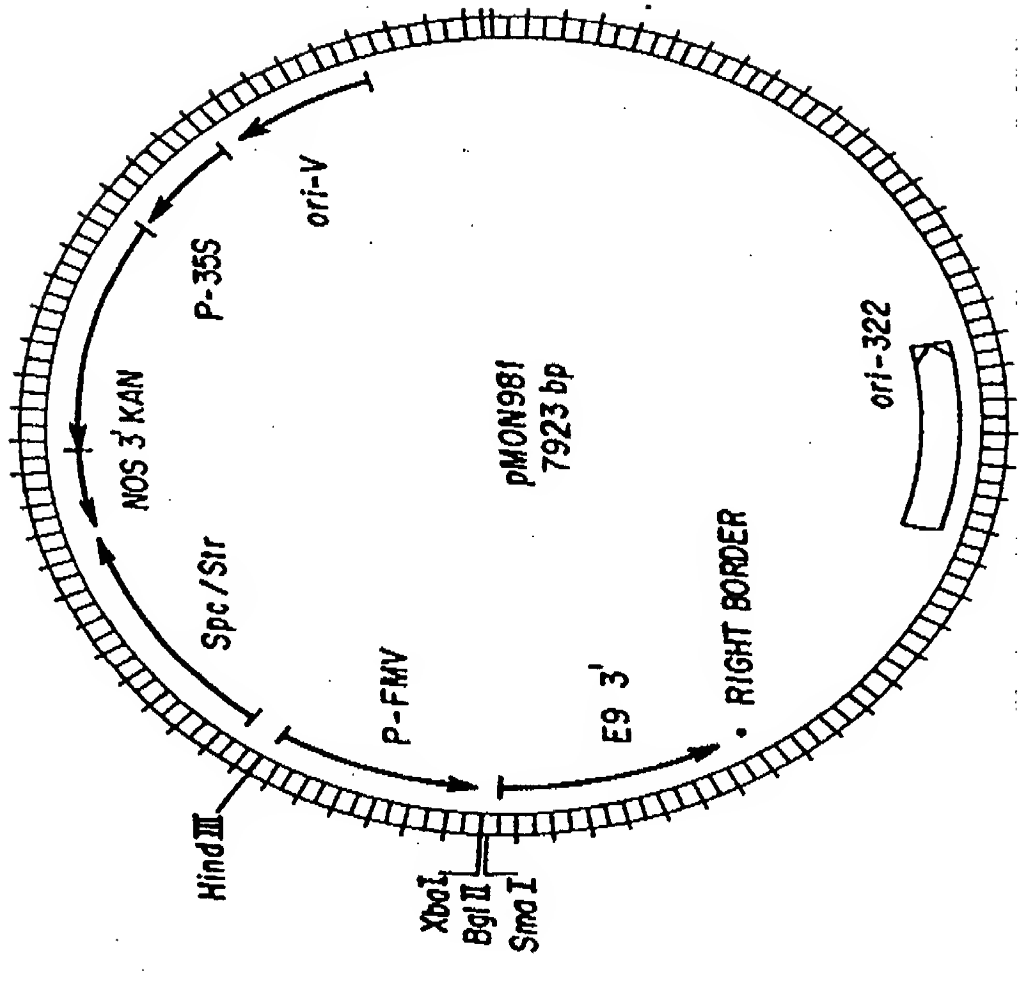

ori-V
P-35S
NOS 3' KAN
Spc/Str
P-FMV
E9 3'
RIGHT BORDER
ori-322
pMON981
7923 bp
HindIII
XbaI
BglII
SmaI

FIG. 5

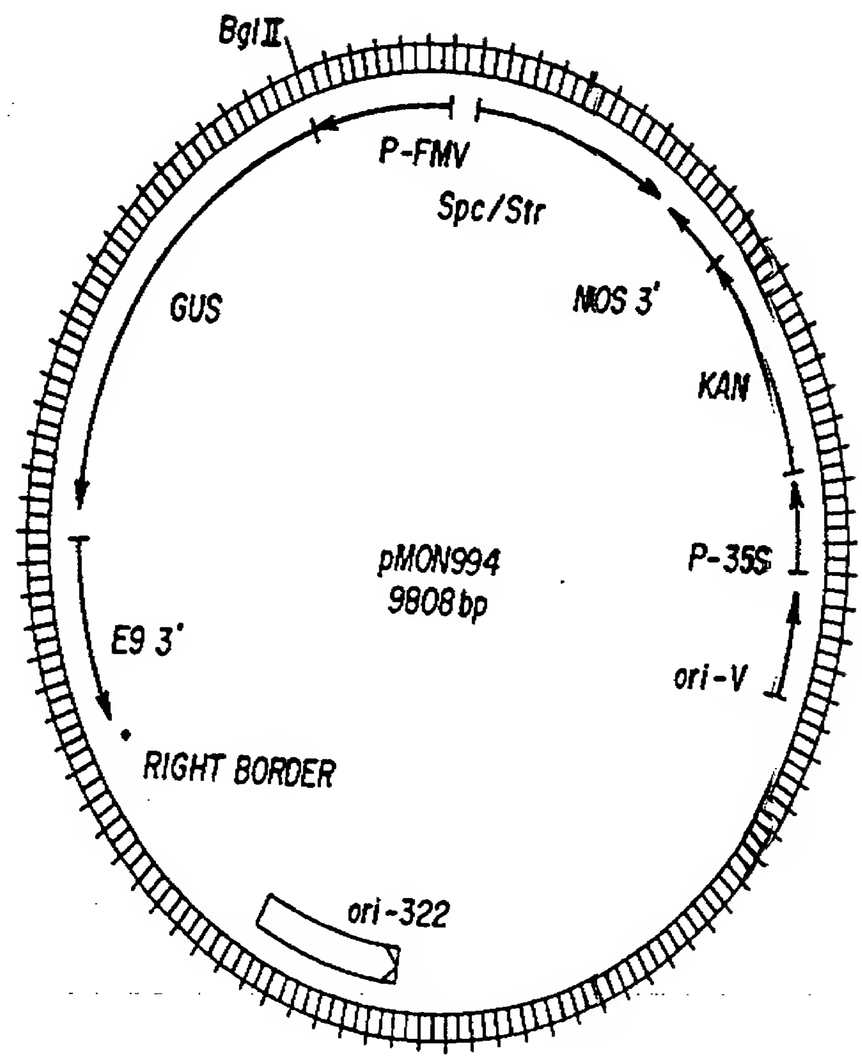

BglII
P-FMV
Spc/Str
GUS
NOS 3'
KAN
pMON994
9808 bp
P-35S
E9 3'
ori-V
RIGHT BORDER
ori-322

FIG. 6

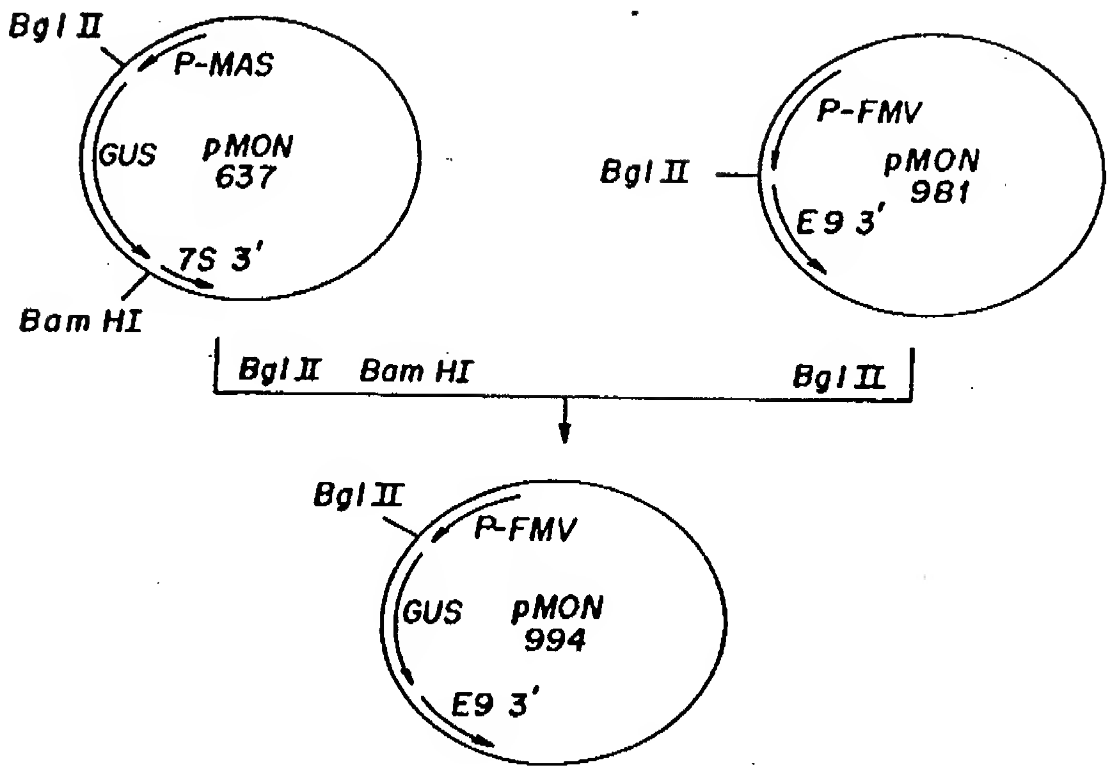

Bgl II
P-MAS
GUS
pMON 637
7S 3'
Bam HI
P-FMV
Bgl II
pMON 981
E9 3'
Bgl II
Bam HI
Bgl II
Bgl II
P-FMV
GUS
pMON 994
E9 3'

FIG. 7

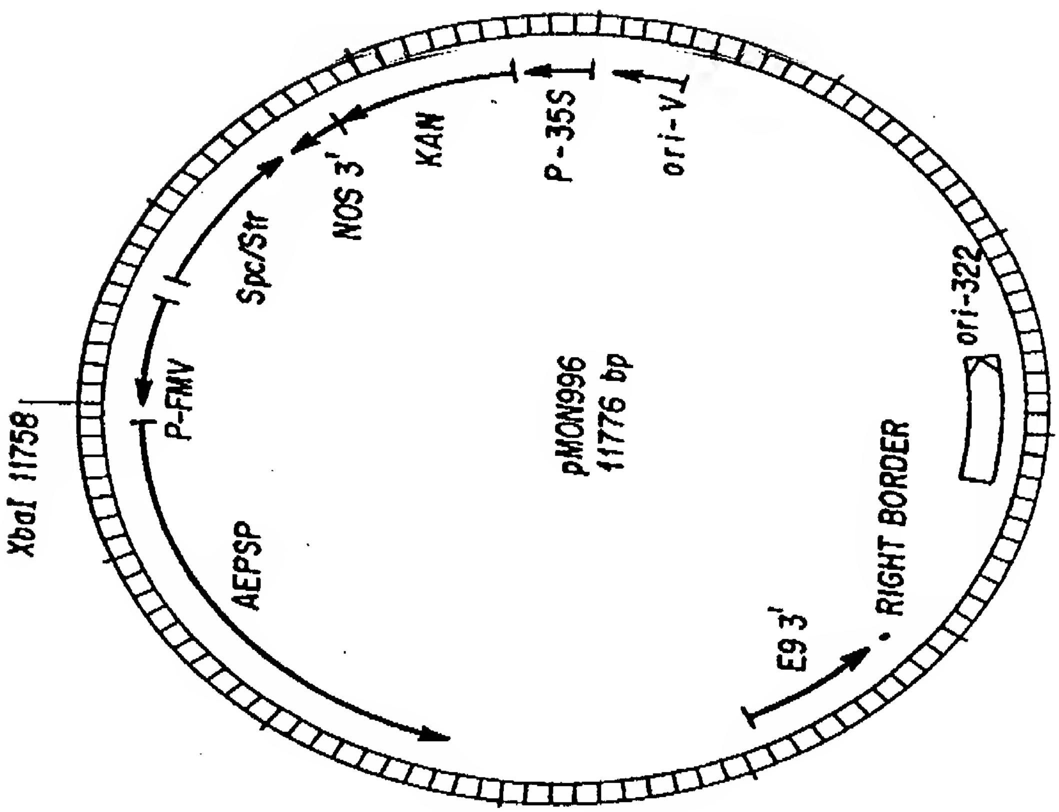
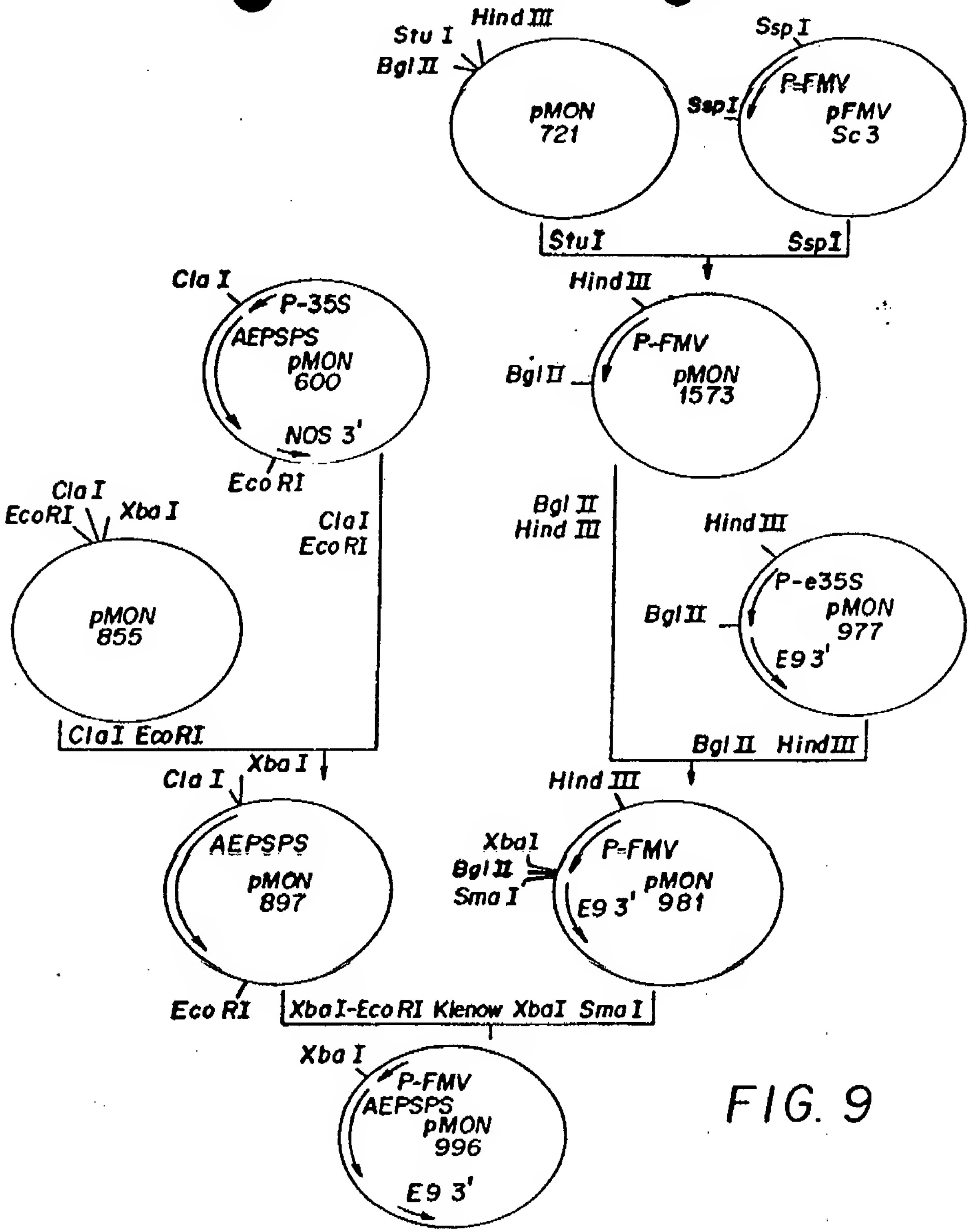
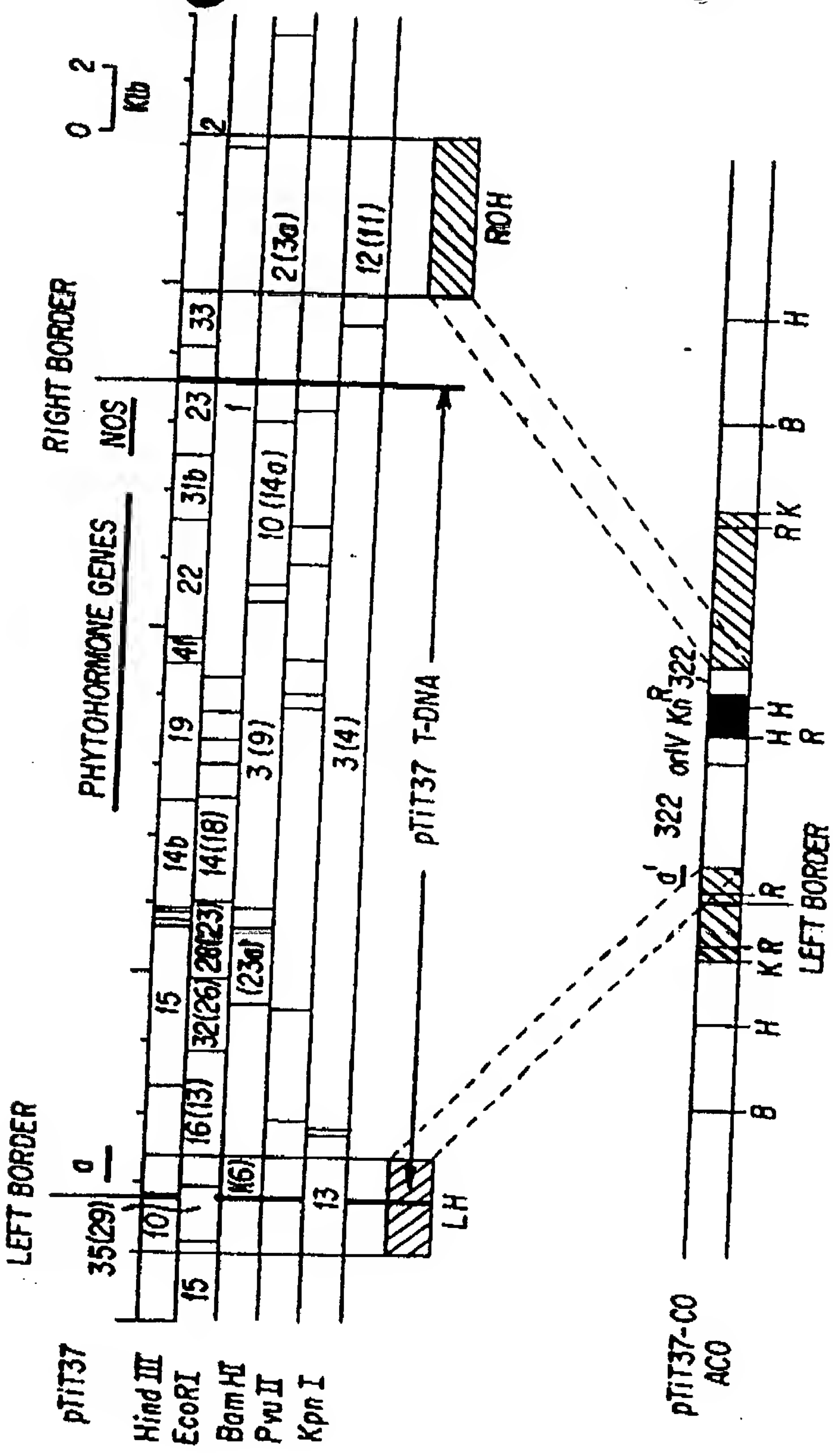

XbaI 11758
P-FMV
Spc/Str
NOS 3'
KAN
P-35S
ori-V
ori-322
pMON996
11776 bp
AEPSP
E9 3'
RIGHT BORDER

FIG. 8



Stu I
Hind III
Bgl II
pMON 721
Ssp I
P-FMV
Ssp I
pFMV Sc 3
Stu I
Ssp I
Cla I
P-35S
AEPSPS
pMON 600
NOS 3'
Eco RI
Hind III
P-FMV
Bgl II
pMON 1573
Cla I
EcoRI
Xba I
pMON 855
Cla I
Eco RI
Bgl II
Hind III
Hind III
P-e35S
Bgl II
pMON 977
E9 3'
Cla I EcoRI
Bgl II Hind III
Cla I
Xba I
AEPSPS
pMON 897
Eco RI
Hind III
XbaI
Bgl II
Sma I
P-FMV
pMON 981
E9 3'
XbaI-Eco RI Klenow XbaI Sma I
Xba I
P-FMV
AEPSPS
pMON 996
E9 3'

FIG. 9



pTiT37
Hind III
EcoRI
BamHI
PvuII
Kpn I
LEFT BORDER
RIGHT BORDER
PHYTOHORMONE GENES
NOS
pTiT37 T-DNA
ROH
LH
0 2 Kb
pTiT37-CO
ACO
322 oriV Kn^R 322
LEFT BORDER

FIG. 10

FIG.11b

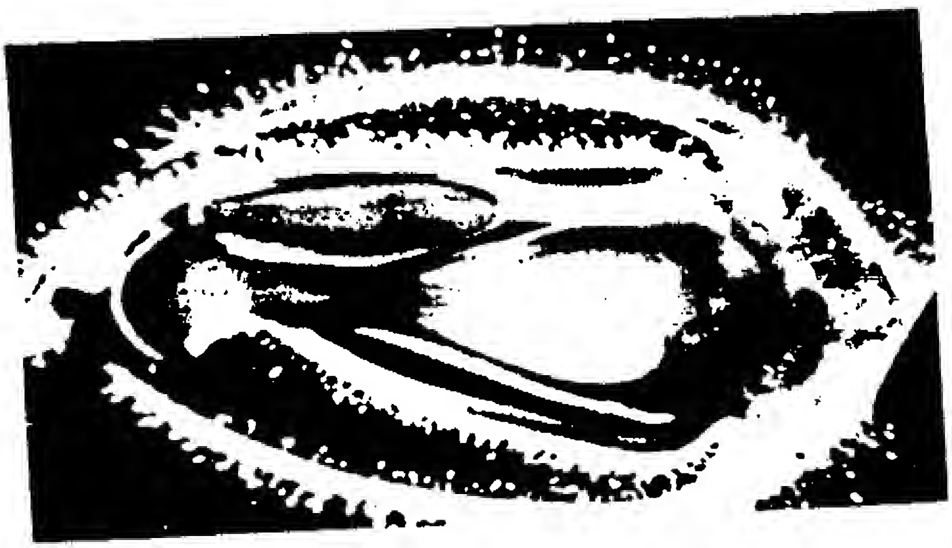
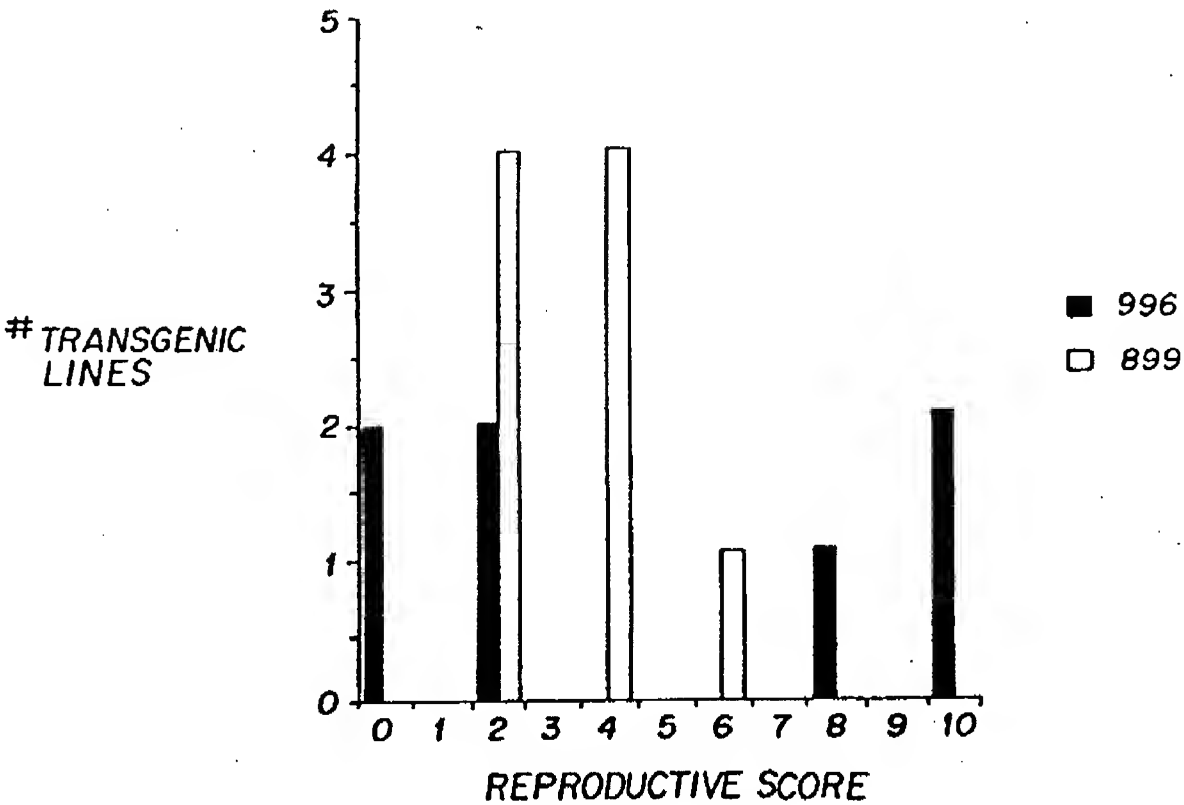


FIG.11a



5
4
3
2
1
0
TRANSGENIC LINES
0
1
2
3
4
5
6
7
8
9
10
REPRODUCTIVE SCORE
996
899

FIG. 12

Volume 17 Number 23 1989 Nucleic Acids Research

The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter

A.Hasegawa[1], J.Verver[+], A.Shimada[1], M.Saito[1], R.Goldbach[+], A.Van Kammen[+], K.Miki[1], M.Kameya-Iwaki[2] and T.Hibi*

Laboratory of Applied Microbiology, National Institute of Agrobiological Resources, Tsukuba Science City, Kannondai, Ibaraki 305, [1]Fundamental Research Laboratory, TONEN Co., 1-3-1 Nishi-tsurugaoka, Ohi-machi, Iruma-gun, Saitama-ken 354 and [2]Department of Plant Protection, National Agricultural Research Center, Tsukuba Science City, Kannondai, Ibaraki 305, Japan

Received July 28, 1989; Revised and Accepted October 24, 1989

EMBL accession no. X15828

ABSTRACT

The complete nucleotide sequence of an infectious clone of soybean chlorotic mottle virus (SoyCMV) DNA was determined and compared with those of three other caulimoviruses, cauliflower mosaic virus (CaMV), carnation etched ring virus and figwort mosaic virus. The double-stranded DNA genome of SoyCMV (8,175 bp) contained nine open reading frames (ORFs) and one large intergenic region. The primer binding sites, gene organization and size of ORFs were similar to those of the other caulimoviruses, except for ORF I, which was split into ORF Ia and Ib. The amino acid sequences deduced from each ORF showed only short, highly homologous regions in several of the corresponding ORFs of the three other caulimoviruses. A promoter fragment of 378 bp in SoyCMV ORF III showed a strong expression activity, comparable to that of the CaMV 35S promoter, in tobacco mesophyll protoplasts as determined by a β-glucuronidase assay using electrotransfection. The fragment contained CAAT and TATA boxes but no transcriptional enhancer signal as reported for the CaMV 35S promoter. Instead, it had sequences homologous to a part of the translational enhancer signal reported for the 5'-leader sequence of tobacco mosaic virus RNA.

INTRODUCTION

Soybean chlorotic mottle virus (SoyCMV) is a member of the caulimovirus group of plant viruses [1]. The virus has been found only in Japan and exclusively in soybean. SoyCMV particles are spherical, about 50 nm in diameter [2] and contain a single molecule of circular double-stranded DNA [3,4]. They occur in electron-dense inclusion bodies in the cytoplasm [2]. The virus is readily transmitted by mechanical inoculation to a few other species of Leguminosae, while the natural vector remains unknown [2]. It is not serologically related to cauliflower mosaic virus (CaMV), carnation etched ring virus (CERV) or peanut chlorotic leaf streak virus [1].

The viral DNA has been cloned and mapped with restriction endonucleases. These data showed that SoyCMV DNA has a size of

approximately 8.2 kbp and contains three single-stranded discontinuities (gaps) typical of the caulimoviruses, one (G1) in the α (-) strand and two (G2 and G3) in the other (+) strand [4]. Three potential primer binding sites for reverse transcription of the viral DNA, a 12-nucleotide sequence complementary to the 3'-terminus of $tRNA_1^{met}$ and two purine-rich sequences, were also detected near the gap sites. However, the physical map and hybridization analyses showed that SoyCMV DNA had no significant sequence homology with CaMV DNA [4].

In this paper we report the complete nucleotide sequence of SoyCMV and compare this sequence with those of three other caulimoviruses, CaMV, CERV and figwort mosaic virus (FMV). In addition, we demonstrate that one of the promoter fragments of SoyCMV DNA had a strong expression activity comparable to that of the CaMV 35S promoter and discuss the structural differences between the two promoters.

MATERIALS AND METHODS

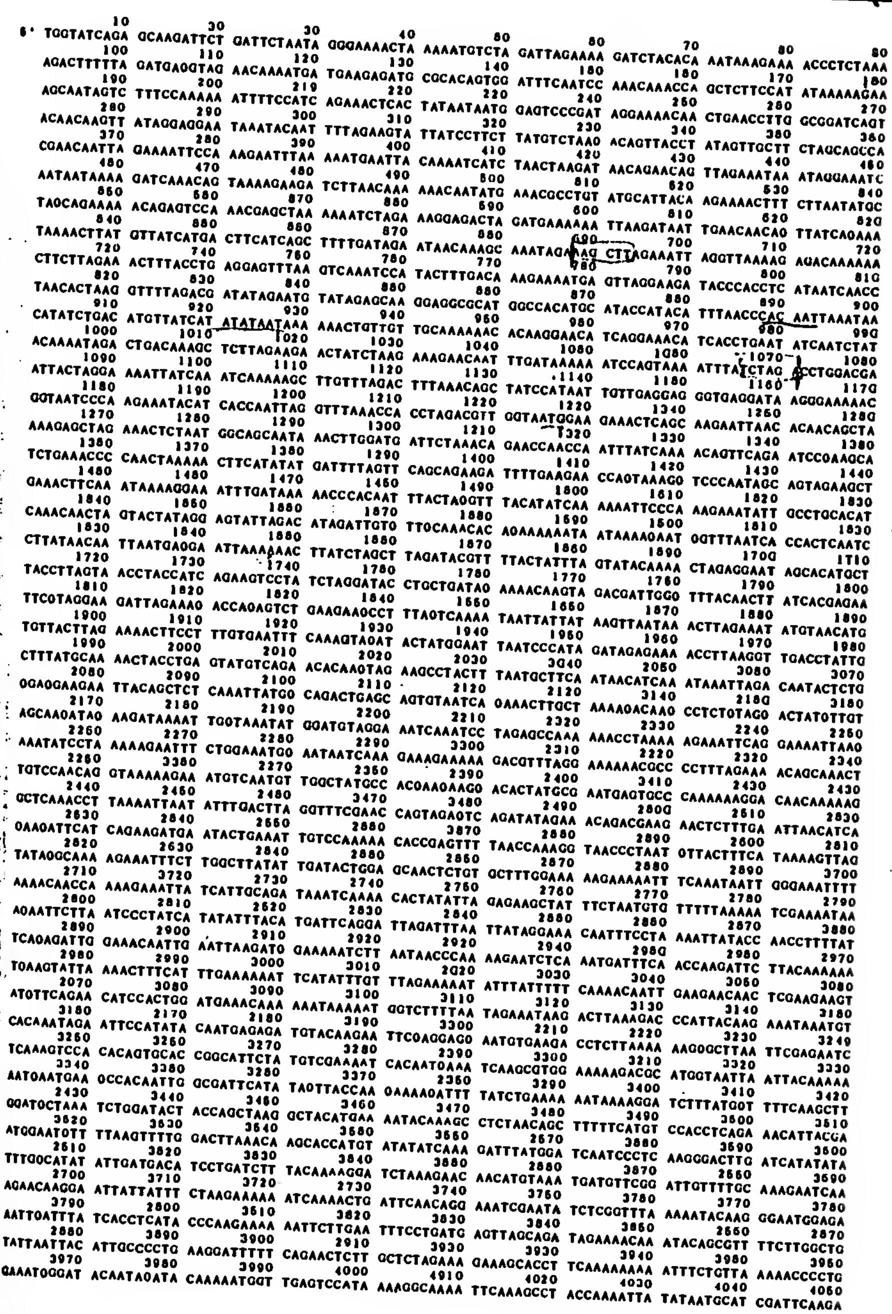
DNA sequencing

A full-length infectious clone of SoyCMV DNA (8.2 kbp) inserted into pSSac [4] was excised with SacI and digested with EcoRI, HindIII and XbaI, respectively. Each restriction fragment was subcloned into pUC118 or pUC119 [5]. These subclones were stepwise deleted by exonuclease III digestion [6,7] and single-stranded DNA templates were prepared for sequence determination [5]. DNA sequencing was carried out according to the dideoxy chain termination method [8] with either *E. coli* DNA polymerase Klenow fragment or Sequenase (United States Biochemicals). Computer analyses of nucleotide and amino acid sequences were performed on an NEC PC9801VM computer using a SDC-GENETIX program (version 6.12).

Construction of a GUS assay plasmid containing SoyCMV promoter IV

A 378 bp HindIII-XbaI fragment of ORF III (nucleotide no. 689-1066), named promoter IV fragment and containing a TATA box at approximately 300 nucleotides upstream of the first ATG of ORF IV, was excised from the full-length viral fragment. This fragment was inserted 5' proximal to the β-glucuronidase (GUS)

TATA ≈ 230



Nucleic Acids Research

```
      4000       4070       4080       4060       4100       4110       4120       4130       4140
CTTTTTAATA GTCGAGACAG ATGCATCGCA ACACTCCTGG AGTGGATGTT TGCGAGCTTT ACCAAGGGAA AGCAAAAAAT CGGACTCGAT
      4160       4160       4170       4180       4100       4200       4210       4220       4236
GAATTCGGGA TACCGACCAT GTGACCTCTG CACAGGTAGC AGTTCAGCTT CAAGCGATAA TTCGCCAGCT GAGATTGACA AATGTCATTC
      4240       4250       4260       4270       4280       4760       4300       4310       4370
AGCCAGTAAA CAGGACACTC ATGTGGCCAG TAAAATAAAG AAACTCGAAA ACGAGCTTCT ACTTTGCAAA TATGTTTCAG GTACCTTCAC
      4320       4340       4280       4260       4370       4280       4300       4400       4410
AGATACGGAA ACAAGATACC CTATAGCAGA ACTGGAGGTT CTTGCTGGAG TAAAAGTCCT AGAAAAATGG AGAATCTACC TCCTACAACC
      4420       4420       4440       4460       4460       4470       4480       4400       4600
GAGGTTCCTC CTCCGCACTG ACAGAAAGTA CTTTGCAGGT TTTTGTAGGT ACAACATCAA GACAGACTAC CGGAACGACA GTCTAACTCA
      4610       4620       4630       4640       4660       4660       4670       4660       4690
GGTGGCAATA CGGTTACAAG CCTATCAACC GTACGTGGAA TTAATCAAAT CAGAAAATAA CCCATTCGAG ATACGCTTAC GCGAGAATGG
      4600       4610       4620       4630       4640       4680       4660       4670       4660
AGCAAGCCAT CAAGCAGTTA AGGAACCAGC AAAAGCAGCT AGATCAGCAG ATCGAGCAGC AAAAACAAAC TCTCAACTCG CTCATTTCGA
      4660       4700       4710       4720       4730       4740       4750       4760       4770
GGAGGGAGGA AGTTACAAAA GGAATCCAAA CATTGGAACT TTTATCTGGA ACAACATTGG TACCAGATCA ACAAATTCAT GAAGCACTGC
      4760       4790       4800       4610       4820       4830       4640       4660       4660
AACCAGCAGT ATCAGAATCA ACAAAAGGCA ACCAAGAAAT AAAAACTTGG AAGCAGATAT TTGATGAAGA GCACGAGCTC ACTCGAGACA
      4870       4860       4860       4900       4610       4920       4620       4640       4660
TGACAAGATC GACGCCACAG TTGGTCAGAA GAAAAAGGCT TACGTCATCT TCGATGGTCC TTGGAAGGGA ATCTACCAAG ATTGGCACAT
      4660       4670       4680       4060       5000       5010       5020       5030       5040
TGTCAAATCA AACGGTAAAT GCCCAACCAT ACAGGTACAA GGGATACAAC TCCCTGGACG AAGCTAAGCT AGCACACAAG CAGGCTTATG
      5060       0060       5070       5060       5060       5100       5110       5170       5120
CTGAAGTAAC TAAGGCAGAC GAAGTAAAGG TTGAAAAAAC TATGAAGTCA TTTGCCAAGA ATAATGTAGC CGAGAAAATT CACGGACTAC
      5140       5160       5160       5170       5160       5190       5200       5310       5220
ACAAGTCAGG ACCAAAAGAA CTGACAGAAG CAGAGTTCTA CCCTAACTGG AAAATGATTA CGGAATGGAC AGAAGAATCA GCCAATCTAG
      5230       0240       5200       5260       5270       5280       5260       5200       5310
GTTTCTACCC AGATTGCAGC AAGCAAGTGA AGGCAGTCTT CTTCATCGGA GCAGATCCTC ACTTACTTAG TTCATTCTAC CAAAGTGGAC
      5320       5330       5340       5360       5360       5370       5360       5390       5466
AGTTAGCTA CATCTATCTA CAAGAAGATG AAGGACAGAA AGGAGCAATC TCAAAGGCTA CCAGCCAGCT CCCAAAAGAA TTAAGAAGGA
      5410       5420       5430       5440       5460       5460       5470       5480       5490
CATGTCAACA ATACCAATTG TCCTTTGCCA AGACAAGAGA ATTCTACTTG GCCACTCAGT CGACCTACCC AGTATTCGAC GAAGAAAAGA
      5600       5610       5620       5630       5640       5650       5600       5570       5550
TGCTTGTCCC AGCAAAGCAT CTGGTTAAGC TAGGCATCTC AGCTAGTTCA TACCCAGAAA ATAAGATCGT AAAAACAAAT TTTAATTTTA
      5560       5600       5610       5620       5630       5640       5660       5660       5670
GTTTGTTTAT TAATTCAGTT GACAAATTGT ACAATTATAT CAGGCAATAT GGAACCACAA TCAAAGGGTT CAAGGTCCTG ATGAAAACAC
      5660       5090       5700       5710       5770       5730       5740       5760       5750
AGCTTTGTCT AGCAGTATGC TTAATCCGCG ACCAAGCTGA AGAAAGCTCA AAAATCATGG TGATGGAGTT CGAACTAGAT ATCTCCACTC
      5770       5780       5700       5800       5810       5820       5830       5840       5850
TAACAGGGAT ATTCAGTAAC CTGCCAAAGG AACTGAAGAA AGCCACATGC GAAAAAATGC ATCGCTACAA GAGCCACTTG TGCGAATCAT
      5860       5870       5600       5660       6000       6010       6620       6930       5646
GCGAGATAAA TTTCCCGGAA CTCTCGGAAA CCATGAACGT CGGCAATGAC GAAAAAAGAA GCACTAAAAG CGTCAGCAGT GACGAAATAA
      5660       5660       5670       5900       5660       6000       6010       0020       6620
ATTTGTCGGC AGAAAACGAC GGATATCAGC ACAGCTAGAT GTAGGACATG CCAGCTGTAA AAGAAACCTC ACCTACTAAT ATGTGGTAGG
      6040       6060       6000       6070       6060       6060       0100       6110       6120
GGACGCTTTA CTTTATTAAA AGTGGTTGGT CAGTAATAAT GTAAGACCCC ACTTCTTTTC TTTTGCTTGC ACGCGAAGGA TGCCGCTCTG
      6130       0140       0180       0160       6170       6160       6160       6200       6210
CCCAGTTGTT AAGGCACCTA TCGCATTATA AATAAGAGAC CAAGGACTCT ATTGTTCCTT GGAGTTTGAT TGAGTAAGGA ATATAGCCAA
      6220       6230       6240       0260       6260       6270       6260       6360       6300
TAGTGCCGTG TAAGGCCAAG TGCTTTTATC CATTTACACT CACTCCCAGT CGGTGGTTTA AAAACCTGGA CCGGCAAAGT CGAGAGACTC
      6210       6320       0230       6340       0360       0360       0270       6360       6266
TAAATTAGAA AAGGAGAAGT CCTTTATACT ATCAAACAAG GAGAGATCCT AAATCTTAAA CACAAAATCC TTTATGAATA AGAAATTGTT
      6400       6410       6430       6430       6440       0460       6460       6470       6466
CCAGCAACTA CCAAGTCTTA AAAAGACCCA GGAAGCAAAA GCAAAGCAAG AACAAGCACA AGAGAAAAGA AAGGTTAGTA AATTATGAAC
      0460       6600       6610       6620       6630       6640       0660       6660       6676
TCAGACTCAG TTATTTGTTT AGCTAGTCTT AGCAATTATT GTCAGTATCA TATATTAGGC ATAATAAACA ATAGGATAGT TCCTATTTTA
      6660       6660       6000       6610       6620       6030       6040       6060       6660
ATAGATACCG GCGCAAGTTG GTCACACATA TCTGCTAGTT TTCTAAAAAC TCATCAAATA AAAAATTGTG AAGAAAAATC TGTTAGAAGA
      6670       6060       0690       6700       6710       0720       6730       6740       6750
TTTGATGGAA CAACTAAAAA ACTAAACAAG AAAACTCTAA TTGAGATAGA CTTGTCGGGT ATACAAAACG TAAAATTAGA ACTTTACGTA
      0760       6770       6760       0760       6600       6810       6020       6830       6846
GATGAAGAAC CAAACAAAAT TTTGCTAGGA ACAGACTTCT TAGAAAATTT TTATTTCAAA ATAGAAAGCG ATTGCCTCTT TCTTAATCAA
      0660       6660       0670       6660       6660       6600       0010       6620       6626
AACCAACATC CACGGTTTAA ACTTCATGAA GATAAAATCT TTGAAATAAT TGAGGATTTA ACTATTCCAA ACGGAATATA GGATGGAATA
      6940       6950       6660       6070       6950       6960       7600       7010       7020
CCAGTTGTAA ACAACACCAC AATGGAAATA GTAGAATTAA AAGATGATAA TCAAGAGTAT TTCTTAGATG CTCTATTAGG AAAAGAAATA
      7030       7040       7050       7060       7070       7660       7090       7100       7110
GAAAAAACCG ATTTTTCAAT TACTGAAAAA GAAAACTTTA AACAAAATAA GTTTAAAGAA TTAAGAAATG TTTTCTCAAG AGATAACATG
      7120       7130       7140       7150       7160       7170       7160       7160       7200
TTAAAATTTG GACTAATGAC AGGAGAAGTT CAAATCCCAA TAGAACAAAC AGATGGTAGT GTTTTCTTAG CCACCATCAG TAAAGAACAA
      7210       7220       7230       7240       7250       7260       7270       7260       7250
ATTACTAAGA GAATTAGTAA AATTGAAGAA AAACAAAGAA GATTAATAAG ATATGTTCAT ATCTCAACTC TTCAAGTTTT AATTAAAAGT
      7360       7210       7220       7330       7340       7260       7360       7370       7288
ACCTTTTTAA AAGGCCTAGA TACACCTCTA GAACTTACAC TTAGAGATAA TAGACTTCTT AATCTAGAAG AATCTAAGAT AGCCGTAGGA
      7290       7400       7410       7420       7420       7440       7460       7460       7478
CACGGAAATC TAAAATATGG AAAAATGAAA TTTGACGTTA ATCTACAATT AGGATTATCA TTAAAAGACT TAGATTTAGA TAGATCAATA
      7460       7490       7600       7610       7620       7630       7640       7656       7660
ATTCTGAACT ATAAATTTCT AAGAAGAAAT TTCATGAAAG AAGGAAACCA CGCGTTTTCA ATAAGTTACA GAATTAACTA TGCTTTGAGT
      7570       7560       7090       7600       7610       7620       7630       7640       7658
AATTCTCACC ATAGCGTAGA ATTTAAACAA AAAGAGAAAA TCTATATTGA TGAACTCTTT TCAGAAGTTT TGGAATTAAA ACACCCTGTG
      7600       7670       7060       7660       7700       7710       7720       7720       7748
TTTTCAAAAC TAACAAAAAG TCAATCTTTA AGAATAGAAC CTTCACCGGT TTTCGAAAAA CCTTTAATTA GTTTTAAAGA AAATCAAAAA
      7750       7760       7770       7760       7760       7600       7610       7620       7628
ACAGAAGAAA AAACTGTTTT TAAACCCCCG AAAAGAGATT TCGAACTAAC TGAAACCTCT AAACTAAAAT CAATGATATC AGATTTAACT
      7640       7660       7860       7670       7660       7660       7600       7010       7510
CAGAAAGTAG TTAATTTAGA TAAAAAGATT TAAGAAATGC CAGGAGATAA CTGGACAAAC TCAATAATCT TTAGAATTAA CACTCTTAAA
      7930       7640       7960       7660       7070       7660       7660       8000       8010
GACCAAATAA TTGAATGGAG TAAAAATCCA ACTGAAGAAA ATGAATATTT TTCAGAAAAA TATTTAGAAA AAATAAACAC TTTAACAAAA
      8020       8020       8040       8060       6060       6070       8060       0090       8100
ACTTTAGATT GGTGTGAAAA GTCCAATAAA TCTGAAGTGG AATTAAGGTG GCTTAACCAA GCATTAAGTC TTCCGACCAA GTGCCACCAG
      8110       8120       8130       8140       8150       8160       8170
CTAGATATTC CTACATACGT AGCGAATCTA GTAGAAATAA TCTGCGCAAC TCTGCCAGAA ACCAACCCCA AAACC 3'
```


gene of pBI 221 (Clontech Laboratories) [9], from which an approx. 800 bp HindIII-XbaI CaMV 35S promoter fragment had been removed previously. The resulted chimeric plasmid was named pBI 241.

Transient expression assay by electrotransfection

Transfection of tobacco mesophyll protoplasts (*Nicotiana tabacum* L. cv. Xanthi NN) with the GUS assay plasmid was carried out according to the procedures for electrotransfection with TMV-RNA [10], with the following modifications. A total of 1 x 10^6 protoplasts was suspended at approximately 5 x 10^5/ml in 0.5 M mannitol with 100 µM $MgCl_2$ and mixed with pBI 241 DNA at 20 to 40 µg/ml and carrier salmon sperm DNA at 50 µg/ml at 4°C. The protoplast suspension was exposed to five square DC pulses of 50 µs duration at 800 V/cm in the electric field at a flow rate of 1.2 x 10^6 protoplasts/min at 4°C by a JASCO CET-100 continuous-flow electrotransfector. For comparison the plasmid pBI 221, containing the CaMV 35S promoter instead of the SoyCMV promoter IV, was also used. The transfected protoplasts were incubated for 36 hours at 25°C as described previously [11]. After incubation the protoplasts were collected by centrifugation, lysed with a Vortex mixer and prepared for the fluorometric enzyme assay [9]. GUS activity was measured using Hitachi F-4010 fluorescence spectrophotometer (exitation 365 nm, emission 455 nm) with a slit width of 1.5 nm.

RESULTS

DNA sequence

The complete nucleotide sequence of the (+) strand of SoyCMV DNA is shown in Figure 1. The numbering begins at the 5' end of the putative (-) strand primer binding site (5'-TGGTATCAGAGC-3') near the G1 site. The G2 (nucleotide no. approx. 4700) and the G3 (nucleotide no. approx. 1170) sites are located in the vicinities of the two purine-rich sequences (5'-GAGGAGGG-3') according to our previous results [4]. The genome comprises 8,175 bp and has a GC content of 34.0%. Hence, it is longer than those of CaMV (8,016-8,032 bp) [12-14], CERV (7,932 bp) [15] and figwort mosaic virus (FMV: 7,743 bp) [16], while the GC content

Fig. 1. The nucleotide sequence of SoyCMV DNA ((+) strand).

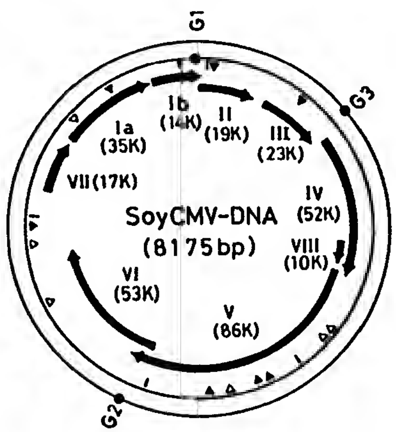


Fig. 2. The organization of the SoyCMV genome. The inner and outer circles represent (-) and (+) strands, respectively, with the gap sites G1, G2 and G3. The arrows depict the potential ORFs in the three reading frames along with the molecular weight in kilodaltons of the deduced translation products. The black and white triangles indicate the positions of potential TATA (TATAT/AAT/A) and TATA-like (TATTT/AAT/A) boxes, respectively. The small bars indicate the positions of potential transcriptional enhancer signals (GTGGT/AT/AT/A).

is lower than in the other caulimoviruses (CaMV, 40.0-40.2%; CERV, 36.4%; FMV, 44%).

Coding regions

Computer analysis of the (+) strand predicts nine putative open reading frames (ORFs) for proteins of more than 10 kDa (Figure 2). Between ORFs VI and VII a large intergenic region (500 nucleotides) is present, while there is no intergenic region between ORF V and ORF VI. No ORF corresponding to more than 10 kDa is present in the (-) strand. The gene organization and the size of the ORFs resemble those of CaMV, CERV and FMV except that ORF I is split into two smaller ORFs, ORF Ia and Ib, and that ORF Ia starts upstream of G1 instead of downstream. Details of the nine ORFs in SoyCMV DNA are given in Table 1. As with CaMV [13] and CERV [15], ORFs Ia-VI and VIII have high lysine contents (molar ratios of 12.5, 10.0, 11.6, 11.4, 11.3, 10.1, 9.5 and 16.5%, respectively). ORF VII is characterized by a high isoleucine content (13.4%) and ORF III contains less proline (6.7%) than CERV ORF III [15].

Table 1. Open reading frames in SoyCMV DNA

ORFs	Nucleotides			Amino acids	
	Start[a]	Stop[a]	Length	Length	Mol. wt
Ia	6,952	7,861	909	303	35,495
Ib	7,867	49	357	119	13,958
II	44	533	489	163	18,754
III	647	1,223	576	192	22,543
IV	1,225	2,545	1,320	440	51,603
V	2,537	4,760	2,223	741	85,753
VI	4,587	5,976	1,389	463	53,013
VII	6,475	6,919	444	148	17,071
VIII	2,189	2,441	252	84	9,654

[a]Nucleotide no., assuming that translation starts at the first in-frame ATG and terminates at the first in-frame stop codon within each open reading frame.

Table 2. Amino acid homologies between SoyCMV and three other caulimoviruses

ORFs	Direct amino acid homologies(%)		
	CaMV	CERV	FMV
I[a]	26.6(282)	30.7(280)	30.9(220)
II	-	-	16.4(67)
III	21.8(55)	24.2(33)	-
IV	24.3(436)	25.1(422)	25.9(417)
V	37.6(575)	35.4(522)	39.6(538)
VI	16.9(278)	43.8(16)	-
VII	-		-
VIII	-		

[a]ORFs I of CaMV, CERV and FMV were compared with ORF Ia of SoyCMV.
The values represent the percentages of maximum matching within the homologous regions indicated as amino acid nos. in parentheses.
- indicates no significant homology found.
The data of CaMV [12], CERV [15] and FMV [16] were used for comparison.

Comparison of amino acid sequences

The amino acid sequences of SoyCMV ORFs Ia-VIII were compared with those of CaMV (isolate Cabb S) ORFs I-VIII [12], CERV ORFs 1-6 [15] and FMV ORFs I-IX [16]. The direct homologies between complete ORFs of SoyCMV and the respective ORFs of the other caulimoviruses were very low (less than 40%) as shown in Table 2. However, some short regions in several ORFs showed high levels of homology.

```
                                     134
N
TMV     92 --LVDKRMERADEATLGSYYTAAAKKRFQFKVVPNYAITTQDAMKN-------------------------  208
SoyCMV 127 --LRDNRLLNLEESKIAVGHGNLKYGKMKFDVNLQLGLSLKDLDLD 170 192 ----PSISYRINYALSNSHHS--  233
CaMV   152 --LIDDRINSRRDCLLGAAKGNLAYGKFMFTVYPKFGISLNTQRLN 195 217 -----MTITYVVGYALNNSHHS--  225
CERV   144 --LIDDRITDRQDSILGAAHGNLVYGKFMFTVYPKYTTSILDQRLD 187 209 -----FSITYLVAYALANSHHS--  230
FMV    149 --LIDDRIVNRKDSLLGAARGNLAYGKFMFMVYPKFALSLQSKNLD 192 214 -----FTVTYLIGYALTNSHHS--
                                                                                   C
```

Fig. 3. Homologous regions in the ORF Ia of SoyCMV and ORFs I of three other caulimoviruses and in the TMV 30K protein. The open boxes indicate identical sequences and the shadowed boxes the sequences conserved among four caulimoviruses.


```
N
       214                           230 373                             C
SoyCMV   --LINLEICNMCYLENFLC---------CPTGKKKCQCWLCHEEGHYANECP--  396
       250                           266 405
CaMV     --MTKLQLCDICYLEEFTC---------CPKGKKDCRCWICNIEGHYANECP--  428
       255                           271 411
CERV     --LINLQLCDICSLQSFFC---------CPKGKKTCRCWVCNIEGHYANECP--  434
       244                           260 402
FMV      --LIKSQLCNICELDNFTC---------CPQGRKKCRCWIQTEEGHYANECP--  425
```

Fig. 4. Homologous regions in the ORFs IV of four different caulimoviruses. The open boxes indicate identical sequences among four caulimoviruses and the shadowed box shows the RNA binding domain.

SoyCMV ORF Ia had two short regions (amino acid no. 145-156 and 200-208) homologous to the middle domain of ORFs I of the other caulimoviruses, where two sequences of twelve and nine amino acids had eight identical amino acids (GNLKYGKMKFDV and YALSNSHHS; underlinings denoting nonhomologous amino acids, see also Figure 3). The 12 residues-long conserved sequence is located in the "transport domain" predicted for CaMV and CERV [15], but the second 9 residues-long conserved sequence resides outside this domain. No other homologous regions were found either in SoyCMV ORF Ia or Ib. The putative molecular weight of the SoyCMV ORF Ia product (35 kDa) is similar to those of the ORF I products of CaMV, CERV and FMV (37, 36, and 37 kDa, respectively). ORF Ia and Ib were separated by a single in-frame stop codon.

ORF IV of SoyCMV showed a short region (amino acid no. 382-395) in which 11 out of 14 amino acids in the sequence (CWLCHEEGHYANEC) were identical to the "RNA binding domain" of the other caulimoviruses [15-18] (Figure 4). This domain has the arrangement of $CX_2CX_4HX_4C$, is referred to as the Cys motif or the zinc-finger, and is known to be a conserved sequence in the *gag* proteins of retroviruses as well as in the coat proteins of CaMV, CERV and FMV. Another limited homologous region was found at approx. 160 residues upstream of this Cys motif (amino acid no. 220-230) where five out of 11 amino acids (CNMCYLENFLC) were conserved (Figure 4). Furthermore, also the "Lys-rich core" (42% of lysine content in amino acid no. 332-379) and the "Glu+Asp rich entities" (23% of glutamic and asparatic acids content in

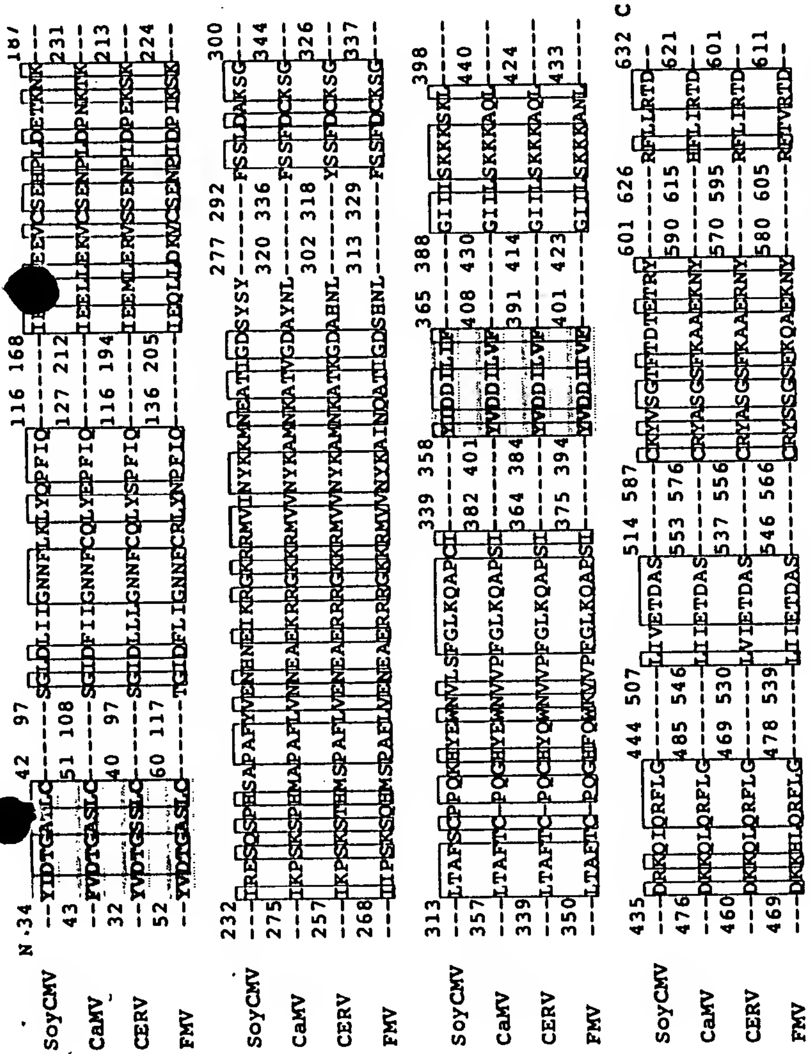


Fig. 5. Homologous regions in the ORFs V of four different caulimoviruses. The open boxes indicate identical sequences among four caulimoviruses, the two shadowed boxes show the protease domain and the reverse transcriptase domain, respectively.


```
         N                          C
         229                      250
SoyCMV   --ADPHLLSSFYQSGLISYIYLQE--
         287                      308
CaMV     --ANPQMVREAYAAGLIKTIYPSN--
         278                      299
CERV     --SDPEGILESFKAGLVRFIYPST--
         279                      300
FMV      --AEPYLVHTAFRAGLAKVLYPSP--
```

Fig. 6. Homologous regions in the ORFs VI of four different caulimoviruses. The open boxes indicate identical sequences among four caulimoviruses. The underlining shows the position reported as a part of the highly conserved sequence among CaMV, CERV and FMV [16].

amino acid no. 2-62 and 47% in amino acid no. 411-439) were detected.

ORF V contained many highly homologous regions to the corresponding ORFs of the other caulimoviruses (Figure 5). There were two characteristic regions, one near the N-terminus (amino acid no. 34-42) and another in the middle (amino acid no. 358-365), in which five out of 9 amino acids in the sequence (YIDTGATLC) were identical to the "protease domain", and five out of 8 amino acids (YIDDILIF) to the "reverse transcriptase domain" of the other caulimoviruses, respectively [15-17].

ORF VI had a short limited homologous region (amino acid no. 241-247) to a part of the "highly conserved sequence" reported for the other caulimoviruses [16] where four out of 7 amino acids in the sequence (GLISYIY) were identical to the other viruses (Figure 6).

ORF Ib, II, III, VII and VIII were also compared with all ORFs of the other caulimoviruses, but no significant homology was found.

Noncoding region and putative promoter regions

The large, 500 bases-long noncoding region between ORF VI and VII was compared with the analogous regions of the other caulimoviruses. A TATA box (TATAAAT) and a TATA-like box (TATTAAA) existed at nucleotide no. 6147 and 6044, respectively. The sequence homology around these boxes to the "highly conserved sequence" of 25 nucleotides [16] was on the average 51% (Figure 7). A 37 bases-long sequence (nucleotide no.

```
        5'
        6031                                     6055
SoyCMV      GGACGCTTTACT--TTATTAAAA-GTGG----
        6134                                     6158
SoyCMV      GCACCTATCGCA--TTATAAATA-AGAG----
        7391                                     7415
CaMV       AAGACCC-TTCCT--CTATATAA-GGAAG----
        7037                                     7063
CERV        AGATCC-TCCCTGGCTATATAAAGGGAG----
        6884                                     6906
FMV        AAGAAT--TCCCT--CTATATAA--GAAG----
                                                          3'
        6346                                                 6382
SoyCMV  ----ATCCTAAATCTTAAACACAAAATCCTT-TATGAATAAG----
        7733                                                 7767
CaMV    ----GAAGTACCGCTTAGGCA-GGAGGCCGT-TAGGGA-AAA----
        7345                                                 7379
CERV    ----GA-GTACCGCCGAGGCA-GGAGGCCGTATAGGGA-AAA----
        7226                                                 7260
FMV     ----GAAGTACCGCCGAGGCA-GGAGGCCGT-TAGGGA-AAA----
```

Fig. 7. Comparative nucleotide sequences around TATA boxes and "bowl sequences" in the large noncoding regions of four caulimoviruses. The open boxes indicate identical sequences and the shadowed boxes TATA (TATAT/AAT/A) or TATA-like (TATTAAA) boxes.

6346-6382) at approx. 190-290 nucleotides downstream of the adjacent TATA and TATA-like boxes showed 46 % of homology to the other "highly conserved sequence" of 35 nucleotides, which folds into a bowl-shaped secondary structure [16,19].

Potential promoter regions, conforming to the consensus sequence TATA/TT/AAT/A are shown in Figure 2. One of these putative promoter regions, promoter IV, showed an actual activity as described below.

Gene expression activity of promoter IV

A GUS assay plasmid, pBI 241, constructed by operon fusion of SoyCMV promoter IV fragment and GUS gene as described in Materials and Methods, exhibited strong, transient gene expression in electrotransfected tobacco protoplasts after 36 hours of incubation (Figure 8). The expression activity was comparable to that of pBI 221 containing the CaMV 35S promoter instead of the SoyCMV promoter fragment. Since no transcriptional or translational products of SoyCMV have been reported so far, there is no information yet whether the promoter IV is active in the viral DNA or not. It is tempting to assume that this promoter is involved in the transcription of both ORF IV and ORF V since other putative signals are lacking in the vicinity upstream of the first ATG of ORF V. This type of promoter has not been reported so far from three other caulimoviruses. As mentioned above, the expression activity of the SoyCMV promoter IV was similar to that of the CaMV 35S promoter, one of the strongest known promoters active in plants. Therefore, the structural differences between the SoyCMV promoter IV and the CaMV 35S promoter were further compared.

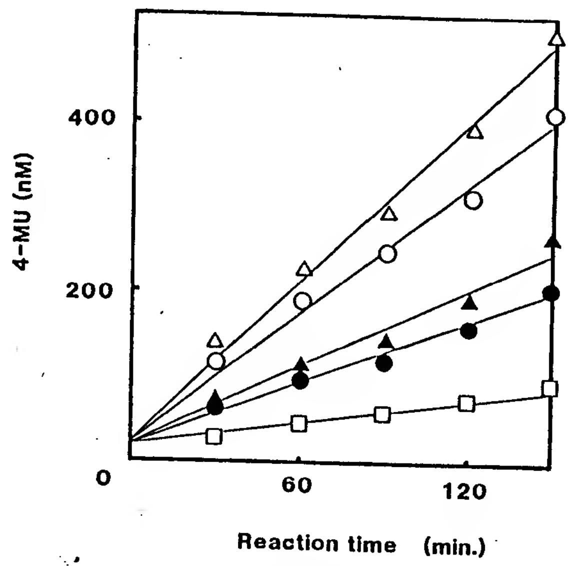


Fig. 8. GUS activity in the extracts of tobacco mesophyll protoplasts transfected with GUS assay plasmids. Tobacco mesophyll protoplasts were electrotransfected with 20 μg/ml (▲) and 40 μg/ml (△) of pBI 241 or with 20 μg/ml (●) and 40 μg/ml (○) of pBI 221, and incubated for 36 hrs. The mock transfection (□) of the protoplasts was carried out without plasmid. The GUS activity of the protoplast extracts was measured according to the methods described in Materials and methods. It is shown as nM of 4-methylumbelliferone (4-MU) produced by the enzyme activity.

Structural differences between the SoyCMV promoter IV and the CaMV 35S promoter

The nucleotide sequences of both promoter fragments are

	689
SoyCMV	5' AGCTTAGAAT TTAGGTTAAA AGAGACAAAA AACTTCTTAG AAACTTTACC TGAGGAGTTT
	7173
CaMV	5' AGTGGAAAAG GAAGGTGGCT CCTACAAATG CCATCATTGC GATAAAGGAA AGGCCATCGT
SoyCMV	AAGTCAAATC CATACTTTGA CAAAGAAAAT GAGTTAGGAA GATACCCACC TCATAATCAA
CaMV	TGAAGATGCC TCTGCCGACA GTGGTCCCAA AGATGGACCC CCACCCACGA GGAGCATCGT
SoyCMV	CCTAACACTA AGGTTTTAGA CGATATAGAA TGTATAGAGC AAGGAGGCCC ATGGCCACAT
CaMV	GGAAAAAGAA GACGTTCCAA CCACGTCTTC AAAGCAAGTG GATTGATGTG ATATCTCCAC
SoyCMV	GCATACCATA CATTTAACCC ACAATTAAAT AACATATCTG ACATGTTATC ATATATAATA
CaMV	TGACGTAAGG GATGACGCAC AATCCCACTA TCCTTCGCAA GACCCTTCCT CTATATAAGG
SoyCMV	AAAAACTGTT GTTGCAAAAA ACACAAGGAA CATCAGGAAA CATCACCTGA ATATCAATCT
CaMV	AAGTTCATTT CATTTGGAGA GGACACGCTG AAATCACCAG TCTCTCTCTA CAAATCTATC
	35S RNA
SoyCMV	ATACAAAATA GACTGACAAA GCTCTTAGAA GAACTATCTA AGAAAGAACA ATTTGATAAA
CaMV	TCTCTCTATA ATAATGTGTG AGTAGTTCCC AGATAAGGGA ATTAGGGTTC TTATAGGGTT
	1066
SoyCMV	AAATCCAGTA AAATTTAT 3'
	7575
CaMV	TCGCTCATGT GTTGAGCATA TAAGAAACCC TTAGTATGTA TTT 3'

depicted in Figure 9. The SoyCMV promoter IV contained CACAAT and TATATAA sequences, as did the CaMV 35S promoter. These sequences were almost equidistant in both promoters. The TATA box of promoter IV overlapped with a poly-A signal, AATAAA, but this signal was inactive in the GUS expression experiment described above. The sequence GTGG is a consensus part of the GTGGA/TA/TA/T sequence (GT motif) which is known as one of the animal transcriptional enhancer cores and found also upstream of the TATA box in the CaMV 35S promoter [20]. However, it was not detected either upstream or downstream of the TATA box of the SoyCMV promoter IV fragment nor in the more upstream region of this fragment in the viral DNA (nucleotide no. 177-688). The TGACG motif which has been identified upstream of histone genes as well as within the CaMV 35S promoter [21] was also lacking in the SoyCMV promoter IV fragment. On the contrary, it was found that there were three characteristic regions which, at the RNA level, were highly homologous to a part of the translational enhancer signal reported for the 5'-leader sequence of tobacco mosaic virus (TMV) RNA [22]. These regions are located approximately 15, 60 and 85 nucleotides downstream of the TATA box in the promoter fragment (Figure 9 and 10). Especially the third region (CUCUUAGAAGAACUAUCUAAGAAAGAA, nucleotide no. 1010-1036) showed 63 % of homology with the 5'-terminal sequence of TMV-RNA (nucleotide no. 4-32 of OM strain [23]). In addition, this sequence is characterized with a GAA repeat, while the corresponding region of TMV has a unique CAA repeat. The cores of these three possible "translational enhancer-like sequences", UUGCAAAA, AUACAAAA and UUAGAAGA, had 75 % of homology to the corresponding TMV sequence. The first of these might have no significance because mRNA will be transcribed from approximately 25-30 nucleotides downstream of the TATA box, as in the case of CaMV 35S promoter. No translational enhancer-like sequences were found in the CaMV 35S promoter fragment.

Fig. 9. Comparative nucleotide sequences of the SoyCMV promoter IV and the CaMV 35S promoter. The open boxes indicate CAAT and TATA boxes. The linear and dotted underlinings show the GT and TGACG motives, respectively. The shadowed boxes indicate regions homologous to the TMV leader sequence.

DISCUSSION

Based on the gene organization, size and putative amino acid homology, each SoyCMV ORF was tentatively assigned to the corresponding ORFs of three other caulimoviruses, except for ORF Ib.

The product of CaMV ORF I is presumed to have important functions including cell-to-cell transport of the virus. It was reported that CaMV ORF I product had some limited homology to the 30K transport protein of TMV [15]. SoyCMV ORF Ia showed a consensus sequence (GNLKYGKMKFDV) for three other caulimoviruses. This sequence and a second, approx. 16 residues upstream, show some limited homology to the TMV 30K protein, as reported for the other caulimoviruses (Figure 3). Another consensus sequence (YALSNSHHS) found in ORFs I appears at approx. 50 residues downstream of the above mentioned sequence. This consensus sequence has no homology with TMV 30K protein.

ORF Ib is unique for SoyCMV and the role of this gene is so far unknown.

CaMV ORF II encodes the aphid transmission factor [24,25]. However, five species of aphids so far examined did not transmit SoyCMV, and the vector of this virus remains unknown [2]. Therefore, the product of SoyCMV ORF II might be inactive for aphid transmission or may have a different role as it shows no significant homology to any of the ORFs of the other caulimoviruses.

SoyCMV ORF IV shows only low amino acid sequence homology to the corresponding ORFs of the other caulimoviruses. This finding fits our previous observation that the virus particle has no serological relationship to that of other caulimoviruses [1]. Still this ORF is assumed to represent the coat protein gene since the derived total protein structure is similar to those of the other caulimoviruses and also contains a consensus Cys motif.

ORF V of SoyCMV exhibits high homology at deduced amino acid level to the corresponding ORFs V of the other caulimoviruses. The presumed protein product of this ORF contains a consensus "protease domain" and a "reverse transcriptase domain" at similar positions as found in ORFs V of the other caulimoviruses. These data strongly suggest that SoyCMV ORF V

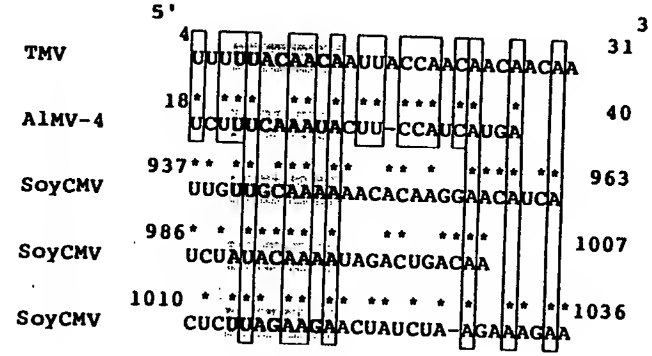


Fig. 10. Homologous regions in the leader sequences of TMV-RNA and AlMV-RNA 4 and in the SoyCMV promoter IV. The asterisks indicate nucleotides also found in the TMV leader sequence. The open boxes indicate identical sequences and the shadowed boxes the possible translational enhancer core-like sequences.

encodes a viral polymerase with reverse transcriptase activity.

CaMV ORF VI has been shown to encode an inclusion body protein [26]. Only limited amino acid sequence homology was detected between this ORF and SoyCMV ORF VI, *i.e.* exclusively in a part of the "highly conserved sequence" [16]. However, based on its size and location in the viral genome, SoyCMV ORF VI is presumed to be an inclusion body gene.

The roles of SoyCMV ORFs III, VII and VIII cannot be speculated upon in the present study.

Detailed analyses on the actual products of the nine ORFs of SoyCMV are in progress.

Potential promoter regions and potential transcriptional enhancer signals were detected in SoyCMV DNA, as shown in Figure 2. There were altogether 22 poly-A signals spread over the genome; therefore their relationship to TATA boxes could not be established. The activities and roles of these potential promoters are not known, except for promoter IV.

If the replication mechanisms of SoyCMV are similar to those of CaMV, there should be a promoter corresponding to the CaMV 35 S promoter, that is involved in the synthesis of the template RNA [27-29]. In the CaMV genome the 35S promoter is located within a large noncoding region upstream of G1 [30]. There were only limited homologies around the TATA (nucleotide no. 6147)

and TATA-like (nucleotide no. 6044) boxes and around "bowl sequsnces" in the large noncoding region of SoyCMV to thoes of the other caulimoviruses. CCACT and GTGGTTT sequences are located at approx. 65 nucleotides upstrsam and approx. 110 nucleotides downstream (not upstream), of the TATA box, respsctively. The TATA-like box has neither CAAT-like sequences within 100 nucleotides upstream nor transcriptional enhancer-like sequences within 200 nucleotides upstream nor downstream. However, in spite of these structural differences, the potential SoyCMV promoter in the noncoding rsgion, especially the TATA box, may correspond to the CaMV 35S promoter, because of its position in the genome and the presence of the downstream "bowl sequence". The transcript analysis are now in progress to prove thie.

Assuming that SoyCMV has a promoter corresponding to the CaMV 19S promoter for the mRNA of ORF VI, it might be located in ORF V at approx. 550 nucleotides upstream of the first ATG of ORF VI. In that case, it will be quite different from the CaMV 19S promoter which is present in a small noncoding region at approx. 45 nucleotides upstream of ORF VI.

Using the GUS assay it was shown that the promoter IV fragmsnt exhibited a strong activity. If this promoter is also active *in vivo*, it might be a promoter for the transcription of ORF IV, the putative coat protein gene, and possibly ORF V, the putative reverse transcriptase gene. Such promoter activity hae not been reported for CaMV and other caulimovirusss, although two promoter-like signals are present in CaMV ORF III [31]. It was proposed that these CaMV promoter-like signale were probably not functional because of their location relative to known RNAs and ORFs [32]. Therefore, the SoyCMV promoter IV may be an extra promoter for the synthesis of other mRNA than the template RNA or the ORF VI mRNA.

In spite of a strong expression activity, comparable to that of the CaMV 35S promoter, the SoyCMV promoter IV dose not contain any known transcriptional enhancer eignals (GT motif), but, instead, contained two or three short homologous ssquences to a part of the TMV RNA leader sequence (Figure 10). A fragment of 67 nucleotides (Ω'; nucleotide no. 2-68) of the noncoding 5'-leader sequence of TMV RNA has been shown to enhance the translation of contiguous foreign gene transcripts *in vitro* and *in vivo* both in sukaryotes and prokaryotes [21]. The sequence of the first half of this fragment is conservsd among the leader sequences of several TMV strains. Enhanced translation was also reported for a 37-nucleotide fragment (nucleotide no. 1-37) of the noncoding 5'-leader sequence of alfalfa mosaic virus (AlMV) RNA 4 that encodes the coat protein [33]. Thie AlMV fragment also has a sequence homologous to part of the Ω' fragmsnt at the same position indicated for the SoyCMV promoter IV (Figure 10). Therefore it is possible that this homologous sequence might be one of the translational enhancer core signals, although some other factors like the secondary structures might relate to the tranelational enhancement. This signal wae also detected downstream of all the other TATA boxes in the SoyCMV genome.

It is not clear in the present study why the SoyCMV promoter IV showed such a strong activity. One possibility ie that another, unidentified transcriptional enhancer signal than the GT motif exists in the promoter fragment. Another possibility is that the transcripts through the promoter IV are efficiently expressed with the aid of translational enhancer-like sequences even if the transcriptional activity of the promoter is not strong enough.

In any case, the data on SoyCMV promoter IV presented here suggest that this promoter may be of a novel type, with functions and structures, which have not been reported for the other caulimoviruses.

Detailed analyss on the actual activities of the possible translational enhancer-like sequences in the SoyCMV promoter IV are now in progrsss as well as analyses on the activities of other potential promoter regions in the genome of this virus.

ACKNOWLEDGEMENTS

We would like to thank Dr. I. Havukkala of the National Institute of Agrobiological Resources for correcting the English language.

*To whom correspondence should be addressed

[1]Department of Molecular Biology and [2]Department of Virology, Agricultural University, De Dreyen 11, 6703 BC Wageningen, The Netherlands

REFERENCES

1. Hibi, T. and Kameya-Iwaki, M. (1988) AAB Descriptions of Plsnt Viruses No. 331.
2. Iwaki, M., Isogawa, Y., Tsuzuki, H. and Honda, Y. (1984) Plant Disease 68, 1009-1011.
3. Hibi, T., Iwaki, M., Saito, Y., Verver, J. and Goldbach, R. (1986) Ann. Phytopath. Soc. Japan 52, 785-792.
4. Verver, J., Schijns, P., Hibi, T. and Goldbach, R. (1987) J. gen. Virol. 68, 159-167.
5. Vieira, J. and Messing. J. (1987) Meth. Enzymol. 153, 3-11.
6. Henikoff, S. (1984) Gene 28, 351-359.
7. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33 103-119.
8. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl Acad. Sci. USA 74, 5463-5467.
9. Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) EMBO J. 6, 3901-3907.
10. Hibi, T., Ksno, H., Sugiura, M., Kazami, T. and Kimura, S. (1988) Plant Cell Reports 7, 153-157.
11. Hibi, T., Kano, H., Sugiura, M., Kazami, T. and Kimura, S. (1986) J. gen. Virol. 67, 2037-2042.
12. Franck, A., Guilley, H., Jonard, G., Richards, K. and Hirth, L. (1980) Cell 21, 285-294.
13. Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R.J. and Messing, J. (1981) Nucleic Acid Res. 9, 2871-2888.
14. Balàzs, E., Guilley, H., Jonard, G. and Richards, K. (1982) Gene 19, 239-249.
15. Hull, R., Sadler, J. and Longstaff, M. (1986) EMBO J. 5, 3083-3090.
16. Richins, R.D., Scholthof, H.B. and Shepherd, J. (1987) Nücleic Ac'd Res. 15, 8451-8466.
17. Fuetterer, J. and Hohn, T. (1987) Trends Biochem. Sci. 12, 92-95.
18. Covey, S.N. (1986) Nucleic Acid Res. 14, 623-633.
19. Füetterer, J., Gordon, K., Bonneville, J.M., Sanfacon, H., Pisan, B., Penswick, J. and Hohn, T. (1988) Nucleic Acid Res. 16 17, 8377-8390.
20. Odell, J.T., Nagy, F. and Chua, N-H. (1985) Nature 313, 810-812.
21. Fsng, R-X., Nagy, F., Sivasubramsniam, S. and Chua, N-H. (1989) Plant Cell 1, 141-150.
22. Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M.A. (1987) Nucleic Acids Res. 15, 3257-3273.
23. Meshi, T., Ishikawa, M., Takamatsu, N., Ohno, T. and Okada, Y. (1983) FEBS Letters 162, 282-285.
24. Woolston, C.J., Covey, S.N., Penswick, J.R. and Davies, J.W. (1983) Gene 23, 15-23.
25. Armour, S.L., Melcher, U., Pirone, T.P., Lyttle, D.J. and Essenberg, R.C. (1983) Virology 129, 25-30.
26. Xiong, C., Muller, S., Lebeurier, G. and Hirth, L. (1982) EMBO J. 1, 971-976.
27. Pfeiffer, P. and Hohn, T. (1983) Cell 33, 781-789.
28. Maule, A.J. (1985) In Davies, J.W.(ed.), Molecular Plant Virology. CRC Press, Boca Raton, Vol. II, pp. 161-190.
29. Gronenborn, B. (1987) In Hohn, T. and Schell, J.(eds.), Plant DNA Infectious Agents. Springer-Verlag, Wien, pp. 1-29.
30. Guilley, H., Dudley, R.K., Jonsrd, G., Balàzs, E. and Richards, K.E. (1982) Cell 30, 763-773.
31. Hohn, T., Richards, K. and Lebeurier, G. (1982) Curr. Top. Micro. Immunol. 96, 193-236.
32. Covey, S.N. (1985) In Davies, J.W.(ed.), Molecular Plant Virology. CRC Press, Boca Raton, Vol. II, pp. 121-159.
33. Jobling, S.A. and Gehrke, L. (1987) Nature 325, 622-625.